

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds
(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146a-23

Perfect score: 151

Sequence: 1 PKPQKTKRNIIRPDQVKFPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

.al number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	100.0	28	20 AAY06673	NC mosaic protein
2	151	100.0	189	23 AAR71258	HCV type 3 capsid
3	151	100.0	191	17 AAR92968	Hepatitis C virus
4	151	100.0	191	17 AAR92969	Hepatitis C virus
5	151	100.0	191	17 AAR92971	Hepatitis C virus
6	151	100.0	319	17 AAR96547	Hepatitis C virus
7	150	99.3	191	17 AAR92970	Hepatitis C virus
8	147	97.4	470	14 AAR34473	Encoded by Hepatit
9	147	97.4	470	14 AAR34474	Encoded by Hepatit

10	147	97.4	470	14	AAR34475	Encoded by Hepatit
11	146	96.7	28	20	AAY06675	NC mosaic protein
12	146	96.7	3010	15	AAR5417	Blood transfusib
13	146	96.7	3010	23	AAE20477	HCV-S1 full-length
14	145	96.0	40	16	AAR84559	Hepatitis C virus
15	145	96.0	40	16	AAR74562	Hepatitis C virus
16	145	96.0	50	16	AAR74563	Hepatitis C virus
17	145	96.0	82	13	AAR24946	Non-A, Non-B Hepat
18	145	96.0	82	13	AAR25110	Non-A, Non-B Hepat
19	145	96.0	191	17	AAR92956	Hepatitis C virus
20	144	95.4	36	16	AAW06487	CN14 fragment of H
21	144	95.4	38	14	AAR30687	HCV capsid peptide
22	144	95.4	38	14	AAR30688	HCV capsid peptide
23	144	95.4	38	14	AAR30689	HCV capsid peptide
24	144	95.4	38	15	AAR54065	Non-A, non-B hepat
25	144	95.4	43	19	AAW37380	Hepatitis C virus
26	144	95.4	44	19	AAW66083	Hepatitis C virus
27	144	95.4	44	21	AAW94410	Human hepatitis C
28	144	95.4	44	21	AAW94409	Human hepatitis C
29	144	95.4	45	21	AAW29534	HCV core-envelope
30	144	95.4	55	13	AAR29535	HCV core-envelope
31	144	95.4	55	13	AAR29536	HCV core-envelope
32	144	95.4	55	13	AAR29537	HCV core-envelope
33	144	95.4	55	13	AAR29542	HCV core-envelope
34	144	95.4	57	13	AAR25121	Non-A, Non-B Hepat
35	144	95.4	61	13	AAR20770	Peptide VIIIE base
36	144	95.4	61	16	AAR69545	Anti-HCV antibody
37	144	95.4	61	17	AAW15425	Prototype peptide
38	144	95.4	61	18	AAW01865	HCV core protein p
39	144	95.4	66	12	AAR12597	PT-NANB viral stru
40	144	95.4	74	17	AAR9530	Hepatitis C virus
41	144	95.4	78	13	AAR29160	Antigen pHCa101.
42	144	95.4	78	13	AAR29161	Antigen pHCb101.
43	144	95.4	79	14	AAR41422	HCV fragment 1 / I
44	144	95.4	79	14	AAR41423	HCV fragment 2 / I

ALIGNMENTS

RESULT 1

AA06673

ID AAY06673 standard; Protein; 28 AA.

XX AAY06673;
AC AAY06673;

XX 17-JUN-1999 (first entry)
DT 17-JUN-1999 (first entry)

XX NC mosaic protein amino acid fragment A.
DE NC mosaic protein amino acid fragment A.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
KW restriction endonuclease assisted ligation; vaccination.
XX Hepatitis C virus.

OS Hepatitis C virus.
XX Hepatitis C virus.

XX WO9910506-A1.
PN WO9910506-A1.

XX 04-MAR-1999.
PD 04-MAR-1999.

XX 21-AUG-1998; 98WO-US17385.
PF 21-AUG-1998; 98WO-US17385.

XX 25-AUG-1997; 97US-0921887.
PR 25-AUG-1997; 97US-0921887.

XX (USSR) US DEPT HEALTH & HUMAN SERVICES.
PA (USSR) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;
PI Fields HA, Khudyakov YE;

XX WPI; 1999-204671/17.
DR WPI; 1999-204671/17.

XX New mosaic protein, comprising a plurality of homologous antigenic
PT peptides from different genotypes of a species - useful for
PT detecting hepatitis infection in an individual

XX Claim 5; Fig 9; 66pp; English.

XX The invention relates to a mosaic protein, comprising a plurality of

CC homologous antigenic peptides from different genotypes of a species. The

CC antigenic peptides are from nucleocapsid (NC) proteins. A method for

CC synthesizing an artificial gene that encodes the mosaic protein is also

CC provided. The method is designated restriction endonuclease assisted

CC ligation (REAL). The mosaic protein and the artificial mosaic protein

CC are useful for detecting a hepatitis infection in an individual. The

CC mosaic gene and protein is also useful for vaccination against

CC infection, especially hepatitis C. The method of synthesizing the

CC artificial gene and the resulting mosaic protein improve the sensitivity,

CC spectrum of immunoreactivity, and antigen specificity of enzyme

CC immunoassays. This provides improved detection of hepatitis C virus.

CC Sequences AAY06673-683 represent amino acid sequence of each monomer

CC comprising the NC mosaic protein.

XX Sequence 28 AA;

Query Match 100.0%; Score 151; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 1.5e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28

Db 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28

RESULT 2

AAB71258

ID AAB71258 standard; protein; 189 AA.

XX

AC AAB71258;

XX

DT 18-NOV-2002 (first entry)

XX

DE HCV type 3 capsid protein fragment.

XX

KW Capsid protein; attenuated vaccine; virucide; antiinflammatory;

KW hepatotropic; yellow fever; Japanese encephalitis; dengue;

KW classical swine fever; bovine viral diarrhoea; hepatitis C.

XX

OS Hepatitis c virus.

XX

PN WO200266621-A1.

XX

PD 29-AUG-2002.

XX

11-FEB-2002; 2002WO-AT00046.

21-FEB-2001; 2001AT-0000272.

(HEIN/) HEINZ F X.

(MAND/) MANDL C.

Heinz FX, Mandl C;

WPI: 2002-667064/71.

Attenuated flavivirus live vaccine, useful for protection against e.g.

yellow fever, comprises virus with attenuating deletion of amino acids

from the capsid protein -

Disclosure; Fig 2; 30pp; German.

This invention describes a novel attenuated flavivirus live vaccine

comprising a flavivirus mutant that has a deletion of at least 4

consecutive amino acids from the capsid protein, provided that the

C-terminal hydrophobic region is not affected by the deletion. The

vaccine of the invention has virucide, antiinflammatory and hepatotropic

activity. The attenuated vaccine, and similar nucleic acid vaccines that

encode the mutated capsid protein, are useful for protection against a

CC wide range of flavivirus diseases, e.g. yellow fever, Japanese

CC encephalitis, dengue, classical swine fever, bovine viral diarrhoea and

CC hepatitis C. The specified deletion: (i) produces a reliably attenuated

CC virus that does not revert to virulence; (ii) is exactly defined and does

CC not effect immune responses to important proteins; and (iii) can not

CC generate a non-natural virus by recombination. The mutant viruses

CC eliminate the need to produce large amounts of infectious/virulent

CC viruses, and can be produced with less expense. The protective response

CC to flavivirus lasts significantly longer than that to killed vaccines.

CC This sequence represents a fragment of the capsid protein from Hepatitis

CC c virus (HCV) type 1 described in the disclosure of the invention.

XX

SQ Sequence 189 AA;

Query Match 100.0%; Score 151; DB 23; Length 189;

Best Local Similarity 100.0%; Pred. No. 1.1e-13;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28

Db 4 PKPQKTKRNTIRRPQDVKPPGGQIVG 31

RESULT 3

AAR92968

ID AAR92968 standard; Protein; 191 AA.

XX

AC AAR92968;

XX

DT 02-OCT-1996 (first entry)

XX

DE Hepatitis C virus isolate HK10 core protein.

XX

KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

XX

OS Hepatitis C virus.

XX

PN WO9605315-A2.

XX

PD 22-FEB-1996.

XX

PF 15-AUG-1995; 95WO-US10398.

XX

PR 15-AUG-1994; 94US-0290665.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (USSH) US SEC DEPT HEALTH.

XX

PI Bukh J, Miller RH, Purcell RH;

XX

DR WPI: 1996-139709/14.

DR N-PSDB; AAT16642.

XX

PT DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

XX

PS Claim 4; Page 207; 340pp; English.

XX

CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers

CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX

SQ Sequence 191 AA;

Query Match 100.0%; Score 151; DB 17; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.1e-13;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPFGGQIVG 28
 DB 5 PKPQKTKRNTIRRPQDVKFPFGGQIVG 32

RESULT 4

AAR92969
 ID AAR92969 standard; Protein; 191 AA.

XX AC AAR92969;

XX DT 02-OCT-1996 (first entry)

XX DE Hepatitis C virus isolate S52 core protein.

XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX KW hepatitis.

XX KW Hepatitis C virus.

XX WO9605315-A2.

XX PD 22-FEB-1996.

XX PF 15-AUG-1995; 95WO-US10398.

XX PR 15-AUG-1994; 94US-0290665.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

XX DR N-PSDB; AAT16643.

XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection

XX PS Claim 4; Page 208; 340pp; English.

XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

XX SQ Sequence 191 AA;

Query Match 100.0%; Score 151; DB 17; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.le-13;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPFGGQIVG 28

DB 5 PKPQKTKRNTIRRPQDVKFPFGGQIVG 32

RESULT 5

AAR92971

ID AAR92971 standard; Protein; 191 AA.

XX AC AAR92971;

XX DT 02-OCT-1996 (first entry)

XX DE Hepatitis C virus isolate DK12 core protein.

XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

OS Hepatitis C virus.

PN WO9605315-A2.

XX PD 22-FEB-1996.

XX PF 15-AUG-1995; 95WO-US10398.

XX PR 15-AUG-1994; 94US-0290665.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

XX DR N-PSDB; AAT16645.

XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection

XX PS Claim 4; Page 209-210; 340pp; English.

XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

XX SQ Sequence 191 AA;

Query Match 100.0%; Score 151; DB 17; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.le-13;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPFGGQIVG 28

DB 5 PKPQKTKRNTIRRPQDVKFPFGGQIVG 32

RESULT 6

AAR96547

ID AAR96547 standard; peptide; 319 AA.

XX AC AAR96547;

XX DT 10-MAR-1997 (first entry)

XX DE Hepatitis C virus types 7c(8a) isolates VN4 amino acids 1-317.

XX KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

FT Misc-difference 144

FT /label= Met, Leu

FT Misc-difference 144..149

FT /label= Val, Ala, Glu, Gly

FT Misc-difference 156

FT /label= Met, Thr, Lys, Arg

FT Misc-difference 157

FT /label= Val, Ala, Asp, Gly

FT Misc-difference 161

FT /label= Gly

FT /note= "amino acid in this position is designated X in
 the specification, but codon usage shows that the

```

FT Misc-difference 167      only possible amino acid at this pos. is Gly"
FT /label= Ser, Arg
FT Misc-difference 167
FT /label= Val, Ala, Glu, Gly
FT Misc-difference 171
FT /label= Gly
FT /note= "amino acid in this position is designated x in
FT the specification, but codon usage shows that the
FT only possible amino acid at this pos. is Gly"
FT
FT Misc-difference 172
FT /label= Cys, Arg, Ser, Gly
FT Misc-difference 174
FT /label= Phe, Leu, Ile, Val
FT Misc-difference 177
FT /label= Phe, Leu
FT Misc-difference 232
FT /label= Met, Val
FT Misc-difference 233
FT /label= Asn, Asp
FT
W09613590-A2.
09-MAY-1996.
23-OCT-1995; 95WO-EP04155.
28-JUN-1995; 95EP-0870076.
21-OCT-1994; 94EP-0870166.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Maertens G, Stuyver L;
XX
XX WPI: 1996-251460/25.
XX N-PSDB; AAT27958.
XX
XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:types
XX - used to develop probes and primers for new sub:types and vaccines
XX to prevent and treat infection
XX
XX Claim 25; Fig 3; 150pp; English.
XX
XX The sequences AAR96526-R96578 represent novel sequences isolated from
XX hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
XX 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
XX 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
XX untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
XX genome. This sequence represents amino acids 1-317 from the HCV types 7c
XX and 8a isolates VNA.
XX The new HCV types were isolated from patients with chronic HCV from the
XX Benelux countries, France, Cameroon and Vietnam, because of their
XX aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
XX amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
XX sequenced either directly or partially and used to classify the new
XX viruses into (sub)types based on comparison with known sequences.
XX The sequences were used to generate the peptides AAR96424-R96524. The
XX sequences can also be used to synthesise probes and primers for the
XX detection of HCV in a sample. The polypeptides can be used to detect
XX anti-HCV antibodies, for HCV typing or to prevent HCV infections.
XX
XX Sequence 319 AA;
XX
XX Query Match 100.0%; Score 151; DB 17; Length 319;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-13;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 PKPQKTKRNTIRRPQDVKEFGGQIVG 28
XX |
XX Db 5 PKPQKTKRNTIRRPQDVKEFGGQIVG 32
XX
XX RESULT 7
XX
XX Query Match 100.0%; Score 151; DB 17; Length 319;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-13;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 PKPQKTKRNTIRRPQDVKEFGGQIVG 28
XX |
XX Db 5 PKPQKTKRNTIRRPQDVKEFGGQIVG 32
XX
XX RESULT 7

```

```

AAR92970
ID AAR92970 standard; Protein; 191 AA.
XX
XX AAR92970;
XX
XX 02-OCT-1996 (first entry)
XX
XX Hepatitis C virus isolate S2 core protein.
XX
XX HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX
XX Hepatitis C virus.
XX
XX W09605315-A2.
XX
XX 22-FEB-1996.
XX
XX 15-AUG-1995; 95WO-US10398.
XX
XX 15-AUG-1994; 94US-0290665.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX (USSH ) US SEC DEPT HEALTH.
XX
XX Bukh J, Miller RH, Purcell RH;
XX
XX WPI: 1996-139709/14.
XX N-PSDB; AAT16644.
XX
XX DNA and amino acid sequence of HCV envelope 1 and core proteins -
XX used to determine HCV genotype and as vaccines against HCV infection
XX
XX Claim 4; Page 209; 340pp; English.
XX
XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX isolates. Isolated cDNA sequences are used for the prodn. of primers
XX useful for detecting the presence of HCV in a sample, the primers
XX are also useful for HCV genotyping. Proteins encoded by the cDNAs
XX can be used in vaccines for immunising against HCV infection. The
XX proteins may also be used to detect antibodies against HCV in serum,
XX saliva, lymphocytes or other mononuclear cells. The antibodies may
XX be used in the prevention of HCV infection.
XX
XX Sequence 191 AA;
XX
XX Query Match 99.3%; Score 150; DB 17; Length 191;
XX Best Local Similarity 96.4%; Pred. No. 1.6e-13;
XX Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 PKPQKTKRNTIRRPQDVKEFGGQIVG 28
XX |
XX Db 5 PKPQKTKRNTIRRPQDVKEFGGQIVG 32
XX
XX RESULT 8
XX
XX AAR34473
ID AAR34473 standard; Protein; 470 AA.
XX
XX AAR34473;
XX
XX 30-JUL-1993 (first entry)
XX
XX Encoded by Hepatitis C virus clone JK3-A.
XX
XX HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
XX polymerase chain reaction; diagnostic method.
XX
XX Hepatitis C virus.
XX
XX JF05068562-A.
XX
XX 23-MAR-1993.

```

```

XX 30-MAY-1991; 91JP-0153736.
XX 30-MAY-1991; 91JP-0153736.
XX (SANW ) SANWA KAGAKU KENKYUSHO CO.
XX WPI; 1993-130638/16.
XX N-PSDB; AAQ40431.
XX DNA and cDNA of hepatitis C virus - useful as probes for
XX diagnosing HCV infection
XX Claim 4; Page 26-28; 44pp; Japanese.
XX cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B
XX (9405 nucleotides long) and 14 shorter clones were isolated by PCR
XX amplification, including clone JK3-A. Primer/probes derived from the
XX sequences of these clones can be used in diagnostic assays for HCV.
XX See AAQ40425-Q40439.
XX Sequence 470 AA;
XX Query Match 97.4%; Score 147; DB 14; Length 470;
XX Best Local Similarity 96.4%; Pred. No. 1.1e-12;
XX Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
XX |||||:|||||:|||||:|||||:|||||
XX 5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32
XX
XX RESULT 9
XX AAR34474
XX ID AAR34474 standard; Protein; 470 AA.
XX AC AAR34474;
XX 30-JUL-1993 (first entry)
XX Encoded by Hepatitis C virus clone JK3-B.
XX HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
XX polymerase chain reaction; diagnostic method.
XX Hepatitis C virus.
XX JP05068562-A.
XX
XX 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
XX |||||:|||||:|||||:|||||:|||||
XX 5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32
XX
XX RESULT 10
XX AAR34475
XX ID AAR34475 standard; Protein; 470 AA.
XX AC AAR34475;
XX 30-JUL-1993 (first entry)
XX Encoded by Hepatitis C virus clone JK3-C.
XX HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
XX polymerase chain reaction; diagnostic method.
XX Hepatitis C virus.
XX JP05068562-A.
XX
XX 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
XX |||||:|||||:|||||:|||||:|||||
XX 5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32
XX
XX RESULT 11
XX AAY06675
XX ID AAY06675 standard; Protein; 28 AA.
XX AC AAY06675;
XX 17-JUN-1999 (first entry)
XX NC mosaic protein amino acid fragment C.
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
XX restriction endonuclease assisted ligation; vaccination.
XX Hepatitis C virus.

```

```

XX Query Match 97.4%; Score 147; DB 14; Length 470;
XX Best Local Similarity 96.4%; Pred. No. 1.1e-12;
XX Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
XX |||||:|||||:|||||:|||||:|||||
XX 5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32
XX
XX RESULT 10
XX AAR34475
XX ID AAR34475 standard; Protein; 470 AA.
XX AC AAR34475;
XX 30-JUL-1993 (first entry)
XX Encoded by Hepatitis C virus clone JK3-C.
XX HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
XX polymerase chain reaction; diagnostic method.
XX Hepatitis C virus.
XX JP05068562-A.
XX
XX 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
XX |||||:|||||:|||||:|||||:|||||
XX 5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32
XX
XX RESULT 11
XX AAY06675
XX ID AAY06675 standard; Protein; 28 AA.
XX AC AAY06675;
XX 17-JUN-1999 (first entry)
XX NC mosaic protein amino acid fragment C.
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
XX restriction endonuclease assisted ligation; vaccination.
XX Hepatitis C virus.

```



```

RESULT 15
AAR74562
ID AAR74562 standard; peptide; 40 AA.
XX
AC AAR74562;
XX
DT 04-JAN-1996 (first entry)
XX
DE Hepatitis C virus core protein antigen peptide.
XX
KW Hepatitis C virus; detection; immunodetection; sensitive.
XX
OS Hepatitis C virus.
XX
PN JP07097398-A.
XX
PD 11-APR-1995.
XX
PF 28-SEP-1993; 93JP-0273938.
XX
28-SEP-1993; 93JP-0273938.
XX
(PA) (ARIM/) ARIMA T.
(PA) (TORA ) TORAY IND INC.
XX
DR WPI; 1995-175362/23.
XX
PT Agent for detection of hepatitis C - also hepatitis C virus (HCV)
PT core protein antigen peptide(s), useful for immuno:detection of HCV
XX
PS Claim 1; Page 7; 10pp; Japanese.
XX
CC AAR74561-63 and AAR74566 are respectively, 24mer, 40mer, 50mer and
CC 30mer antigenic peptides derived from the hepatitis C virus clone
CC S29 genes III and IV. They are useful in the detection of and
CC diagnosis of Hepatitis C virus infection. The peptides are used in a
CC detection method which has a higher sensitivity than conventional
CC detection agents.
XX
SQ Sequence 40 AA;
Query Match 96.0%; Score 145; DB 16; Length 40;
Best Local Similarity 96.4%; Pred. No. 1.6e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 PKPQRTKRTIRRPQDVKEPPGGQIVG 28
Dh 4 PKPQRTKRTIRRPQDVKEPPGGQIVG 31

```

Search completed: August 7, 2003, 11:14:05
Job time : 39.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-23
Perfect score: 151
Sequence: 1 PKQKTKRNTIRRPQDVKFGGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
al number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	151	100.0	114	2 S41359	genome polyprotein
2	151	100.0	114	2 S41358	genome polyprotein
3	151	100.0	124	2 S41360	genome polyprotein
4	151	100.0	411	2 PC2061	genome polyprotein
5	151	100.0	492	2 S41288	genome polyprotein
6	147	97.4	123	2 S41361	genome polyprotein
7	147	97.4	782	2 S19875	genome polyprotein
8	146	96.7	108	2 S41356	genome polyprotein
9	146	96.7	411	2 PC2060	genome polyprotein
10	146	96.7	3010	1 A45573	genome polyprotein
11	145	96.0	88	2 S21336	genome polyprotein
12	144	95.4	108	2 S41353	genome polyprotein
13	144	95.4	108	2 S41355	genome polyprotein
14	144	95.4	108	2 S41357	genome polyprotein
15	144	95.4	108	2 S41348	genome polyprotein
16	144	95.4	112	2 S41371	genome polyprotein
17	144	95.4	112	2 S41341	genome polyprotein
18	144	95.4	114	2 S41370	genome polyprotein
19	144	95.4	114	2 S41369	genome polyprotein
20	144	95.4	114	2 S41368	genome polyprotein
21	144	95.4	115	2 S41342	genome polyprotein
22	144	95.4	115	2 S41344	genome polyprotein
23	144	95.4	115	2 S41350	genome polyprotein
24	144	95.4	115	2 S41354	genome polyprotein
25	144	95.4	115	2 S41345	genome polyprotein
26	144	95.4	115	2 S41347	genome polyprotein
27	144	95.4	115	2 S41343	genome polyprotein
28	144	95.4	118	2 S41346	genome polyprotein
29	144	95.4	266	2 PQ0393	genome polyprotein

30	144	95.4	369	2 S21471	genome polyprotein
31	144	95.4	441	2 S12707	genome polyprotein
32	144	95.4	513	2 PC1284	genome polyprotein
33	144	95.4	520	2 JQ1925	polyprotein - hepa
34	144	95.4	523	2 JQ1926	polyprotein - hepa
35	144	95.4	550	2 JH0711	genome polyprotein
36	144	95.4	782	2 S19876	genome polyprotein
37	144	95.4	782	2 S18031	genome polyprotein
38	144	95.4	782	2 S18032	genome polyprotein
39	144	95.4	787	2 PN0677	hypothetical prote
40	144	95.4	874	2 JQ0883	genome polyprotein
41	144	95.4	874	2 JQ0881	genome polyprotein
42	144	95.4	876	2 PC2219	polypeptide - hepa
43	144	95.4	3010	1 GNAVTC	genome polyprotein
44	144	95.4	3010	1 GNAVCI	genome polyprotein
45	144	95.4	3010	1 S18030	genome polyprotein

ALIGNMENTS

RESULT 1

S41359
genome polyprotein - hepatitis C virus (genotype 3, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41359
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41359
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229462; NID:g443886; PIDN:CAA82600.1; PID:g443887
A:Experimental source: genotype 3, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match	100.0%	Score 151;	DB 2;	Length 114;
Best Local Similarity	100.0%	Pred. No. 2.4e-14;		
Matches	28;	Conservative	0;	Mismatches 0;
Indels	0;	Gaps	0;	
Qy	1	PKQKTKRNTIRRPQDVKFGGGQIVG	28	
Db	5	PKQKTKRNTIRRPQDVKFGGGQIVG	32	

RESULT 2

S41358
genome polyprotein - hepatitis C virus (genotype 3, N1) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41358
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41358
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229461; NID:g443884; PIDN:CAA82599.1; PID:g443885
A:Experimental source: genotype 3, N1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match	100.0%	Score 151;	DB 2;	Length 114;
-------------	--------	------------	-------	-------------

Best Local Similarity 100.0%; Pred. No. 2.4e-14; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
|||||
DB 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 3

S41360 genome polyprotein - hepatitis C virus (genotype 3, N3) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41360
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994.
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

Accession: S41360
Molecule type: genomic RNA
Residues: 1-124 <VAN>
A:Cross-references: EMBL:Z29463; NID:9443888; PIDN:CAA82601.1; PID:9443889

A:Experimental source: genotype 3, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-124/Product: core protein #status predicted <MAT>

Query Match 100.0%; Score 151; DB 2; Length 124;

Best Local Similarity 100.0%; Pred. No. 2.6e-14; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
|||||
DB 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 4

PC2061 genome polyprotein N2 - hepatitis C virus

N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2061
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France.
A:Reference number: PC2060; MUID:94197744; PMID:8147893

Accession: PC2061
Molecule type: mRNA
Residues: 1-411 <LIJ>
A:Cross-references: GB:L12355; NID:9410169; PIDN:AAA20155.1; PID:9410170
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: App; capsid protein; envelope protein; glycoprotein; nonstructural protein;
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,334,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 151; DB 2; Length 411;

Best Local Similarity 100.0%; Pred. No. 8.3e-14; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
|||||
DB 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 5

S41288 genome polyprotein - hepatitis C virus (fragment)

N:Contains: core protein; envelope protein; NS1 protein
C:Species: hepatitis C virus

C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41288

R:Seelig, R.

submitted to the EMBL Data Library, December 1993

A:Reference number: S41288

A:Accession: S41288

A:Molecule type: genomic RNA

A:Residues: 1-492 <SEE>

A:Cross-references: EMBL:X76918

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; envelope protein; nonstructural protein;

F:1-191/Product: core protein #status predicted <COR>

F:192-372/Product: envelope protein #status predicted <ENV>

F:373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match 100.0%; Score 151; DB 2; Length 492;

Best Local Similarity 100.0%; Pred. No. 9.9e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28

|||||

DB 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 6

S41361

genome polyprotein - hepatitis C virus (genotype 3, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 3, N4

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41361

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41361

A:Molecule type: genomic RNA

A:Residues: 1-123 <VAN>

A:Cross-references: EMBL:Z29464; NID:9443890; PIDN:CAA82602.1; PID:9443891

A:Experimental source: genotype 3, N4

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-123/Product: core protein #status predicted <MAT>

Query Match 97.4%; Score 147; DB 2; Length 123;

Best Local Similarity 92.9%; Pred. No. 9.5e-14;

Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28

|||||

DB 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 7

S19875

genome polyprotein - hepatitis C virus (isolate JK3) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 prote

C:Species: hepatitis C virus

A:Variety: isolate JK3

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C:Accession: S19875

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Viru

A:Reference number: S18029

A:Accession: S19875

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61592; NID:959482; PIDN:CAA43789.1; PID:959483

A:Experimental source: isolate JK3

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstruct

F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 97.4%; Score 147; DB 2; Length 782;
Best Local Similarity 96.4%; Pred. No. 5.8e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32

RESULT 8
S41356
genome polyprotein - hepatitis C virus (genotype 2, N5) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
'ariety: genotype 2, N5
ate: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
Accession: S41356
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41356
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:CROSS-references: EMBL:229459
A:Experimental source: genotype 2, N5
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 96.7%; Score 146; DB 2; Length 108;
Best Local Similarity 96.4%; Pred. No. 1.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32

RESULT 9
PC2060
genome polyprotein N1 - hepatitis C virus
Contains: envelope protein E1; nonstructural protein E2/NS1
Species: hepatitis C virus
Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
Accession: PC2060
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France.
A:Reference number: PC2060; MUID:94197744; PMID:8147893
A:Accession: PC2060
A:Molecule type: mRNA
A:Residues: 1-411 <LII>
A:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 96.7%; Score 146; DB 2; Length 411;
Best Local Similarity 92.9%; Pred. No. 4.3e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32

RESULT 10
A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hiji
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese car
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:CROSS-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g22161
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: DEXH motif
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NSA>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NSB>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 96.7%; Score 146; DB 1; Length 3010;
Best Local Similarity 96.4%; Pred. No. 3.1e-12;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32

RESULT 11
S21336
genome polyprotein S29 (core protein region) - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
Accession: S21336
R:Sato, A.
submitted to the EMBL Data Library, April 1992
A:Description: A sensitive serodiagnosis of hepatitis C virus infection with two c
A:Reference number: S21336
A:Accession: S21336
A:Molecule type: genomic RNA
A:Residues: 1-88 <SAT>
A:CROSS-references: EMBL:X65548; NID:g59492; PIDN:CAA46517.1; PID:g59493
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 96.0%; Score 145; DB 2; Length 88;
Best Local Similarity 96.4%; Pred. No. 1.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
DB 11 PKPQKTKRNTIRRPQDVKFFGGGQIVG 38

RESULT 12
S41353
genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
N:Contains: core protein

```
C:Species: hepatitis C virus
A:Variety: genotype 2, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41353
R:van Doorn, L.J.: Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41353
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229456
A:Experimental source: genotype 2, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 95.4%; Score 144; DB 2; Length 108;
Best Local Similarity 96.4%; Pred. No. 2.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 PKPQRTKRTNRRPDVRFPGGGQIVG 28
|||||  |||||  |||||  |||||  |||||
5 PKPQRTKRTNRRPDVRFPGGGQIVG 32

Db

RESULT 13
S41355
genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41355
R:van Doorn, L.J.: Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41355
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229458
A:Experimental source: genotype 2, N4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 95.4%; Score 144; DB 2; Length 108;
Best Local Similarity 96.4%; Pred. No. 2.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 PKPQRTKRTNRRPDVRFPGGGQIVG 28
|||||  |||||  |||||  |||||  |||||
5 PKPQRTKRTNRRPDVRFPGGGQIVG 32

Db

RESULT 14
S41357
genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41357
R:van Doorn, L.J.: Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41357
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229460
A:Experimental source: genotype 2, N6
```

```
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 95.4%; Score 144; DB 2; Length 108;
Best Local Similarity 96.4%; Pred. No. 2.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 PKPQRTKRTNRRPDVRFPGGGQIVG 28
|||||  |||||  |||||  |||||  |||||
5 PKPQRTKRTNRRPDVRFPGGGQIVG 32

Db

RESULT 15
S41348
genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41348
R:van Doorn, L.J.: Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41348
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229451
A:Experimental source: genotype 1, N6
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 95.4%; Score 144; DB 2; Length 108;
Best Local Similarity 96.4%; Pred. No. 2.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 PKPQRTKRTNRRPDVRFPGGGQIVG 28
|||||  |||||  |||||  |||||  |||||
5 PKPQRTKRTNRRPDVRFPGGGQIVG 32

Db

Search completed: August 7, 2003, 11:21:46
Job time : 9.54545 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146A-23

Perfect score: 151
Sequence: 1 PKQKTKRNTIRPDVQKPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

al number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	96.7	3010	1	POLG_HCVJT
2	144	95.4	513	1	POLG_HCVJ2
3	144	95.4	520	1	POLG_HCVH4
4	144	95.4	520	1	POLG_HCVHK
5	144	95.4	737	1	POLG_HCVJ5
6	144	95.4	737	1	POLG_HCVJ7
7	144	95.4	3010	1	POLG_HCVBK
8	144	95.4	3010	1	POLG_HCVJA
9	144	95.4	3011	1	POLG_HCVH
10	144	95.4	3033	1	POLG_HCVJ6
11	144	95.4	3033	1	POLG_HCVJ8
12	137	90.7	3010	1	POLG_HCVTM
13	136	90.1	3011	1	POLG_HCVI
14	54	35.8	548	1	SYN_BRUMA
15	52	34.4	794	1	FUR1_HUMAN
16	50.5	33.4	3988	1	POLG_BVDVN
17	49	32.5	282	1	RR4_TOBAC
18	49	32.5	496	1	CAT3_MAIZE
19	48.5	32.1	309	1	FN3X_HUMAN
20	48	31.8	219	1	Y401_BUCAP
21	47	31.1	58	1	RL29_YEAST
22	47	31.1	194	1	RS7_FUGRU
23	47	31.1	194	1	RS7_HUMAN
24	47	31.1	793	1	FUR1_MOUSE
25	46	30.5	556	1	PDPK_HUMAN
26	46	30.5	559	1	PDPK_MOUSE
27	46	30.5	559	1	PDPK_RAT
28	46	30.5	797	1	FUR1_BOVIN
29	46	30.5	1311	1	FMR2_HUMAN
30	45.5	30.1	303	1	YB1_XENLA
31	45.5	30.1	638	1	YAB1_YEAST
32	45.5	30.1	732	1	TAU_MOUSE
33	45.5	30.1	751	1	TAU_RAT

34 45.5 30.1 948 1 CHR1_MOUSE
35 45 29.8 293 1 RK4_SPIOL
36 45 29.8 422 1 PSY_ATH
37 45 29.8 595 1 FRA_DROME
38 45 29.8 941 1 CHR1_XENLA
39 45 29.8 1425 1 MADI_HUMAN
40 44.5 29.5 279 1 Y144_MYCE
41 44.5 29.5 584 1 VATA_METH
42 44.5 29.5 718 1 ICAL_RABIT
43 44 29.1 190 1 RS7_MANSE
44 44 29.1 194 1 RS7_XENLA
45 44 29.1 332 1 PI11_MOUSE

ALIGNMENTS

RESULT 1

POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
"Molecular cloning of hepatitis C virus genome from a single Japanese
carrier: sequence variation within the same individual and among
infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
precursor polyprotein, commonly with Asp or Glu in the P6
position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
[RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; D11168; BAA01943.1;
CC PIR; A45573; A45573.
CC PDB; 1A1Q; 25-MAR-98.
CC PDB; 1JXP; 14-JAN-98.
CC MEROPS; S29.001;
CC MEROPS; U39.001;
CC InterPro; IPR001410; DEAD.

FT	CARBOHYD	418	418
----	----------	-----	-----

1. CARBONATED WATER, : ; (FOUNTAIN),
410 N STARKES (BURNING, : ;)

```

FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 95.4%; Score 144; DB 1; Length 520;
Best Local Similarity 96.4%; Pred. No. 6.7e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTIRPQDVKPPGGQIVG 28
    ||||| ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTIRPQDVKPPGGQIVG 32

RESULT 5
POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC EMBL: D10075; BAA00969.1;
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CC CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 734 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

```

FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 95.4%; Score 144; DB 1; Length 737;
Best Local Similarity 96.4%; Pred. No. 9.8e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTIRPQDVKPPGGQIVG 28
    ||||| ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTIRPQDVKPPGGQIVG 32

RESULT 6
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC EMBL: D10077; BAA00971.1;
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.

```

DR Pfam: PF01542; HCV_core; 1.
 DR PFam: PF01539; HCV_env; 1.
 DR PFam: PF01560; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polypeptide: Glycoprotein; Coat protein; Envelope protein;
 Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 131
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 67DFAE11854122F2 CRC64;
 Query Match 95.4%; Score 144; DB 1; Length 737;
 Best Local Similarity 96.4%; Pred. No. 9.8e-14;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNTIRRPQDVKFGGQIVG 28
 DB 5 PKPQKTKRNTIRRPQDVKFGGQIVG 32
 RESULT 7
 ID POLG.HCVBK STANDARD; PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 NT 01-AUG-1992 (Rel. 23, Last sequence update)
 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS5B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 from human carriers";
 RT J. Virol. 65:1105-1113(1991).
 RL [2]
 RN SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 phosphorylation mediated by CAMP-dependent protein kinase.";
 Eur. J. Biochem. 237:611-618(1996).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parage H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 Moonaw E.W., Adachi T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 trypsin-like fold and a structural zinc binding site.";
 Cell 87:331-342(1996).
 RL [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 Protein Sci. 7:837-847(1998).
 RL [5]
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL: M58335; AAA72945.1;
 DR PIR: A38465; GNVVTC
 DR PDB: 1A1Q; 25-MAR-98.
 DR PDB: 1JXP; 14-JAN-98.
 DR PDB: 1NS3; 08-APR-98.
 DR PDB: 1C2P; 15-NOV-00.
 DR PDB: 1CSJ; 08-NOV-99.
 DR PDB: 1GX5; 09-APR-02.
 DR PDB: 1GX6; 10-APR-02.
 DR PDB: 1QUV; 26-JUN-00.
 DR PDB: 8OHM; 20-APR-99.
 DR MEROPS: S29.001;
 DR MEROPS: U39.001;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.

DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RGRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3010 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT STRAND 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186

FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1204 1204
 FT STRAND 1680 1688
 SO SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;
 Query Match 95.4%; Score 144; DB 1; Length 3010;
 Best Local Similarity 96.4%; Pred. No. 4.7e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
 Db 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32
 RESULT 8
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RP [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=9108850; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RA "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90208; BAA14233.1;
 DR PIR: A39253; GNWVCJ.

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; M67463; AAA45534.1;
CC PUR; A36814; GNMVCH.
CC PDB; 1HE1; 25-NOV-98.
CC PDB; 1A1V; 16-FEB-99.
CC PDB; 1A1R; 17-JUN-98.
CC MEROPS; S29.001; -.
CC MEROPS; U39.001; -.
CC TRANSFAC; T04155; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRP.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00998; Viral_RdRP; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC SMART; SM00487; DEXdc; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

623 623 CARBOHYD
645 645 CARBOHYD
1224 1224 STRAND
1232 1232 TURN
1236 1236 TURN
1239 1239 HELIX
1247 1247 TURN
1251 1255 STRAND
1258 1271 HELIX
1272 1272 TURN
1277 1280 STRAND
1281 1282 TURN
1283 1285 STRAND
1291 1295 STRAND
1296 1301 HELIX
1302 1303 TURN
1312 1316 STRAND
1317 1319 TURN
1323 1335 STRAND
1336 1340 TURN
1343 1347 STRAND
1352 1353 TURN
1360 1361 TURN
1362 1366 STRAND
1368 1368 STRAND
1373 1375 STRAND
1376 1377 TURN
1378 1380 STRAND
1382 1385 HELIX
1389 1393 STRAND
1397 1409 HELIX
1410 1411 TURN
1414 1417 STRAND
1419 1420 TURN
1432 1436 STRAND
1438 1439 TURN
1450 1453 STRAND
1456 1463 STRAND
1471 1478 STRAND
1480 1480 STRAND
1481 1488 HELIX
1489 1490 TURN
1497 1501 STRAND
1507 1507 STRAND
1511 1511 STRAND
1514 1527 HELIX
1532 1544 HELIX
1550 1550 STRAND
1555 1564 STRAND
1570 1578 HELIX
1579 1580 TURN
1584 1597 HELIX
1598 1598 TURN
1606 1611 HELIX
1614 1618 TURN
1622 1623 STRAND
1627 1627 STRAND
1635 1636 STRAND
1640 1652 HELIX
SQ SEQUENCE 3011 AA: 327142 MW: 772CBB29CCD94753 CRC64;

Query Match 95.4%; Score 144; DB 1; Length 3011;
Best Local Similarity 96.4%; Pred. No. 4.7e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 PKPQRTKRNTRRRPDVVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRRRPDVVKFPGGGQIVG 32
|||||
|||||

RESULT 10
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
ID POLG_HCVJ6

```

P26660;
 01-AUG-1992 (Rel. 23, Created)
 01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=9204440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.;
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D00944; BAA00792.1; -
 CC PIR: JQ1303; JQ1303.
 CC HSP: P27958; 1HEI.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV core.
 CC InterPro: IPR002519; HCV env.
 CC InterPro: IPR002531; HCV NS1.
 CC InterPro: IPR002518; HCV NS2.
 CC InterPro: IPR004109; HCV NS3.
 CC InterPro: IPR000745; HCV NS4A.
 CC InterPro: IPR001490; HCV NS4B.
 CC InterPro: IPR002868; HCV NS5A.
 CC InterPro: IPR002166; HCV NS5B.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4A; 1.
 CC Pfam: PF01001; HCV_NS4B; 1.

DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NSL; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191
 FT MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383
 FT MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 733
 FT NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CHAIN 734 1010
 FT NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1011 1619
 FT PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1620 1866
 FT NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1867 2017
 FT NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2018 3033
 FT RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369
 FT POTENTIAL.
 FT ACT_SITE 1087 1087
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1111 1111
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1169 1169
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1234 1241
 FT ATP (POTENTIAL).
 FT SITE 1320 1323
 FT DECH_BOX.
 FT CARBOHYD 196 196
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 209 209
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 234 234
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 305 305
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 417 417
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 423 423
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 430 430
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 448 448
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 477 477
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 534 534
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 542 542
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 558 558
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 578 578
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 627 627
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 649 649
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1091 1091
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 2038 2038
 FT N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 2811 2811
 FT N-LINKED (GLCNAC). (POTENTIAL).
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 95.4%; Score 144; DB 1; Length 3033;
 Best Local Similarity 96.4%; Pred. No. 4.7e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQRKTRNTIRRPQDVKFGGGQIVG 28
 DB 5 PKPQRKTRNTIRRPQDVKFGGGQIVG 32
 RESULT 11
 ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: capsid protein c (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]

DR PIR: A39166; GNVVC3.
 DR PDB: 1AIV; 16-FEB-99.
 DR MEROPS: S29.001; 25-NOV-98.
 DR MEROPS: U39.001; 25-NOV-98.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_core; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 3011 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 90.18; Score 136; DB 1; Length 3011;
 Best Local Similarity 89.38; Pred No. 7.4e-12;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNTIRPQDVKFPGGQIVG 28
 Db 5 PKQKKNKNTNRRPQDVKFPGGQIVG 32
 RESULT 14
 SYN_SYN BRUMA STANDARD; PRT; 548 AA.
 ID AC P10723;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Asparaginyl-tRNA synthetase, cytoplasmic (EC 6.1.1.22) (Asparagine--
 DE tRNA ligase) (AsnRS) (Potentially protective 63 kDa antigen).
 OS Brugia malayi (Filarial nematode worm).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Brugia.
 OX NCBI_TaxID=6279;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88288341; PubMed=2840577;
 RA Perrine K.G., Denker J.A., Nilsen T.W.;
 RT "A multi-copy gene encodes a potentially protective antigen in Brugia
 RT malayi.";
 RL Mol. Biochem. Parasitol. 30:97-104(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217950; PubMed=3368467;
 RA Nilsen T.W., Maroney P.A., Goodwin R.G., Perrine K.G., Denker J.A.,
 RA Nanduri J., Kazura J.W.;
 RT "Cloning and characterization of a potentially protective antigen in
 RT lymphatic filariasis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:3604-3607(1988).
 RN [3]
 RP FUNCTION.
 RX MEDLINE=96049569; PubMed=7589498;
 RA Kron M., Marquard K., Hartlein M., Price S., Leberman R.;
 RT "An immunodominant antigen of Brugia malayi is an asparaginyl-tRNA
 RT synthetase.";
 RL FEBS Lett. 374:122-124(1995).
 CC -1- FUNCTION: POTENTIALLY PROTECTIVE ANTIGEN IN LYMPHATIC FILARIASIS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-asparagine + tRNA(Asn) -> AMP +
 CC diphosphate + L-asparaginyl-tRNA(Asn).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to licensee@isb-sib.ch).
 CC -----
 DR EMBL: J03971; AAA27852.1; -;
 DR EMBL: J03266; AAA27849.1; -;
 DR PIR: A28209; A28209.
 DR PIR: A54510; A54510.
 DR InterPro: IPR004522; AsnS.
 DR InterPro: IPR004364; tRNA-synt_2.
 DR InterPro: IPR002312; tRNA-synt_2.
 DR InterPro: IPR004365; tRNA-anti.
 DR InterPro: IPR006195; tRNA_ligase_II.
 DR Pfam: PF00152; tRNA-synt_2; 1.
 DR Pfam: PF01336; tRNA-anti; 1.
 DR PRINTS: PR01042; TRNASYNTHASP.
 DR TIGRFAMs: TIGR00457; asnS; 1.

FT SITE 70 75 CLEAVAGE (SECOND AUTO-).
 FT SITE 104 107 CLEAVAGE (FIRST AUTO-).
 FT SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 759 762 CELL SURFACE SIGNAL.
 FT SITE 773 779 TRANS GOLGI NETWORK SIGNAL.
 SQ SEQUENCE 794 AA; 86678 MW; 10C44DD3892EF85D.CRC64;

Query Match 34.4%; Score 52; DB 1; Length 794;
 Best Local Similarity 50.0%; Pred. No. 6.8;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 4 ORKTKRNTIRRPQDVKEP 21
 Db 102 KRTKRDVYQEPDKEP 119

Search completed: August 7, 2003, 11:19:55
 Job time : 6.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-23

Perfect score: 151
Sequence: 1 PKPQKTKRNTIRRPQDVKFFGGQIVG 28

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

al number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rviro:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	45	12	Q68307
2	151	100.0	45	12	Q68308
3	151	100.0	45	12	Q68310
4	151	100.0	60	12	Q8JYS2
5	151	100.0	61	12	Q8JYS3
6	151	100.0	73	12	Q8JYR4
7	151	100.0	100	12	Q8QPE5
8	151	100.0	100	12	Q8QPE7
9	151	100.0	100	12	Q8QPE2
10	151	100.0	100	12	Q8QPE1
11	151	100.0	100	12	Q8QPE4
12	151	100.0	100	12	Q8QPE0
13	151	100.0	100	12	Q8QPE7
14	151	100.0	100	12	Q8QPE4
15	151	100.0	100	12	Q8QPE5
16	151	100.0	100	12	Q8QPE9

17 151 100.0 100 12 Q8QPE1 hepatitis c

18 151 100.0 100 12 Q8QPE3 hepatitis c

19 151 100.0 100 12 Q8QPE6 hepatitis c

20 151 100.0 100 12 Q8QPE8 hepatitis c

21 151 100.0 109 12 Q81340 hepatitis c

22 151 100.0 114 12 Q8QPE4 hepatitis c

23 151 100.0 114 12 Q8892 hepatitis c

24 151 100.0 114 12 Q8893 hepatitis c

25 151 100.0 119 12 Q8BCX2 hepatitis c

26 151 100.0 119 12 Q8BCX1 hepatitis c

27 151 100.0 119 12 Q8BCX0 hepatitis c

28 151 100.0 119 12 Q8BCW9 hepatitis c

29 151 100.0 119 12 Q8BCW8 hepatitis c

30 151 100.0 119 12 Q8BCW7 hepatitis c

31 151 100.0 119 12 Q8BCW6 hepatitis c

32 151 100.0 119 12 Q8BCW4 hepatitis c

33 151 100.0 122 12 Q8QRJ1 hepatitis c

34 151 100.0 122 12 Q8QMT6 hepatitis c

35 151 100.0 122 12 Q8QMT8 hepatitis c

36 151 100.0 122 12 Q8QRJ4 hepatitis c

37 151 100.0 122 12 Q8QRJ5 hepatitis c

38 151 100.0 122 12 Q8QMT5 hepatitis c

39 151 100.0 122 12 Q8QMT9 hepatitis c

40 151 100.0 122 12 Q8QMT1 hepatitis c

41 151 100.0 122 12 Q8QMT0 hepatitis c

42 151 100.0 122 12 Q8QRJ7 hepatitis c

43 151 100.0 122 12 Q8QRJ3 hepatitis c

44 151 100.0 122 12 Q8QRJ8 hepatitis c

45 151 100.0 122 12 Q8QRJ9 hepatitis c

ALIGNMENTS

RESULT 1

Q68307 PRELIMINARY; PRT; 45 AA.

AC Q68307;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID-11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-HCV-B21;

RA Songsivilai S., Kanistanon D., Kunkitti R.;

RT "Identification and characterisation of Thai isolates of hepatitis C virus."

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL; U23745; AAA65052.1; .

DR InterPro; IPR002522; HCV_capsid.

DR Pfam; PF01543; HCV_capsid; 1.

KW Polyprotein.

FT NON_TER. 45

SQ SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 45;

Best Local Similarity 100.0%; Pred. No. 1e-15; Mismatches 0; Indels 0; Gaps 0;

Matches 28; Conservative 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGQIVG 28

Db 5 PKPQKTKRNTIRRPQDVKFFGGQIVG 32

```

RESULT 2
Q68308 ID Q68308 PRELIMINARY; PRT; 45 AA.
AC Q68308;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polypeptide) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-BB38;
RA Songsivillai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
RL virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U23748; AAA65053.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 45;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
Db |||||
5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32

RESULT 3
Q68310 ID Q68310 PRELIMINARY; PRT; 45 AA.
AC Q68310;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polypeptide) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-C94009;
RA Songsivillai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
RL virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U23748; AAA65055.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 45;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
Db |||||
5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32

RESULT 4
Q8JYS2 ID Q8JYS2 PRELIMINARY; PRT; 60 AA.
AC Q8JYS2;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIG253;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF506613; AAM33389.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 60 AA; 6725 MW; ACCE7D9C8B90299A CRC64;

Query Match 100.0%; Score 151; DB 12; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.4e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKRTIRRPQDVKFPGGGQIVG 28
Db |||||
5 PKPQRTKRTIRRPQDVKFPGGGQIVG 32

RESULT 5
Q8JYS3 ID Q8JYS3 PRELIMINARY; PRT; 61 AA.
AC Q8JYS3;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIG256;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF506612; AAM33388.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 61 AA; 6881 MW; B92CCE7D9C8B9029 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 45;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Query Match          100.0%; Score 151; DB 12; Length 61;
Best Local Similarity 100.0%; Pred. No. 1.4e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
    |||||
Db 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32
    |||||

RESULT 6
Q8JYR4
ID Q8JYR4 PRELIMINARY; PRT; 73 AA.
AC Q8JYR4
DT 01-JUN-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG291;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF506621; AAM33397.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 1 73
FT NON_TER 73 73
FT SEQUENCE 73 AA; 8201 MW; 5EDC082DAE8CB796 CRC64;

Query Match          100.0%; Score 151; DB 12; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.8e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
    |||||
Db 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
    |||||

MULT 7
JP85
ID Q8QP85 PRELIMINARY; PRT; 100 AA.
AC Q8QP85;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-602;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF070180; AAL58594.1; -.

```

```

DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
FT SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match          100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
    |||||
Db 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32
    |||||

RESULT 8
Q8QP87
ID Q8QP87 PRELIMINARY; PRT; 100 AA.
AC Q8QP87;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-233;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AY070178; AAL58592.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
FT SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match          100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGQIVG 28
    |||||
Db 5 PKPQKTKRNTIRRPQDVKFFGGGQIVG 32
    |||||

RESULT 9
Q8QP72
ID Q8QP72 PRELIMINARY; PRT; 100 AA.
AC Q8QP72;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-609;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified

```

```

RT in St. Petersburg.
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070193; AAL58607.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

--SULT 10
Q8QP71 PRELIMINARY; PRT; 100 AA.
AC Q8QP71;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN SEQUENCE FROM N.A.
RP STRAIN=637;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
RT in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070194; AAL58608.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

RESULT 11
Q8QP74 PRELIMINARY; PRT; 100 AA.
AC Q8QP74;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

RESULT 12
Q8QP80 PRELIMINARY; PRT; 100 AA.
AC Q8QP80;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN SEQUENCE FROM N.A.
RP STRAIN=781;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
RT in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070185; AAL58599.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 100 AA; 11289 MW; F3BC90F5F595F18B CRC64;

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

RESULT 13
Q8QP77 PRELIMINARY; PRT; 100 AA.
AC Q8QP77;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFPGGGQIVG 32

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=748;
RX MEDLINE=21904745; PubMed=11907242;
RA Kallina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: AY070188; AAL58602.1; -.
InterPro: IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 100 AA; 11289 MW; F3BC90F5F595F18B CRC64;

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKRTTIRRPQDVKFPGGQIVG 28
|||||
DB 5 PKPQRTKRTTIRRPQDVKFPGGQIVG 32

RESULT 14
Q8QP84 PRELIMINARY; PRT; 100 AA.
ID Q8QP84;
AC Q8QP84;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=718;
RX MEDLINE=21904745; PubMed=11907242;
RA Kallina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: AY070181; AAL58595.1; -.
InterPro: IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKRTTIRRPQDVKFPGGQIVG 28
|||||
DB 5 PKPQRTKRTTIRRPQDVKFPGGQIVG 32

RESULT 15
Q8QP75 PRELIMINARY; PRT; 100 AA.
ID Q8QP75;
AC Q8QP75;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=680;
RX MEDLINE=21904745; PubMed=11907242;
RA Kallina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: AY070190; AAL58604.1; -.
InterPro: IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 100 AA; 11289 MW; F3BC90F5F595F18B CRC64;

Query Match 100.0%; Score 151; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKRTTIRRPQDVKFPGGQIVG 28
|||||
DB 5 PKPQRTKRTTIRRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 11:18:55
Job time : 28.6364 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-23

Perfect score: 151

Sequence: 1 PKPQKTKRNTIRPDQVKFPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

al number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	28	3	US-08-921-887-23
2	151	100.0	191	2	US-08-290-665A-187
3	151	100.0	191	2	US-08-290-665A-188
4	151	100.0	191	2	US-08-290-665A-190
5	151	100.0	191	5	PCT-US95-10398-187
6	151	100.0	191	5	PCT-US95-10398-188
7	151	100.0	191	5	PCT-US95-10398-190
8	151	100.0	319	3	US-08-836-075A-44
9	151	100.0	319	4	US-08-835-886C-230
10	150	99.3	191	2	US-08-290-665A-189
11	150	99.3	191	5	PCT-US95-10398-189
12	146	96.7	28	3	US-08-921-887-25
13	146	96.7	450	4	US-08-635-886C-191
14	146	96.7	450	4	US-08-635-886C-192
15	145	96.0	191	2	US-08-290-665A-175
16	145	96.0	191	5	PCT-US95-10398-175
17	144	95.4	34	3	US-08-380-160-6
18	144	95.4	43	4	US-09-020-846-36
19	144	95.4	44	3	US-08-380-160-2
20	144	95.4	44	4	US-09-389-756-1
21	144	95.4	45	3	US-08-380-160-1
22	144	95.4	61	1	US-07-946-034-9
23	144	95.4	61	1	US-08-083-947-23
24	144	95.4	61	1	US-08-530-550-3
25	144	95.4	61	1	US-08-262-037-26
26	144	95.4	61	5	PCT-US93-08638-9
27	144	95.4	61	5	PCT-US94-07088-23

28 144 95.4 61 5 PCT-US95-13660-3 Sequence 3, Appl1
29 144 95.4 74 3 US-08-836-075A-10 Sequence 10, Appl1
30 144 95.4 74 4 US-08-635-886C-198 Sequence 198, App
31 144 95.4 100 4 US-08-635-886C-232 Sequence 232, App
32 144 95.4 108 3 US-08-836-075A-14 Sequence 14, Appl1
33 144 95.4 115 1 US-08-324-977-8 Sequence 8, Appl1
34 144 95.4 115 2 US-08-384-616-8 Sequence 8, Appl1
35 144 95.4 115 2 US-08-904-688A-8 Sequence 8, Appl1
36 144 95.4 115 3 US-09-315-850-8 Sequence 8, Appl1
37 144 95.4 123 2 US-08-501-195-2 Sequence 2, Appl1
38 144 95.4 124 1 US-08-244-116B-15 Sequence 15, Appl1
39 144 95.4 137 3 US-08-836-075A-46 Sequence 46, Appl1
40 144 95.4 138 3 US-08-836-075A-60 Sequence 60, Appl1
41 144 95.4 140 2 US-08-501-195-1 Sequence 1, Appl1
42 144 95.4 154 3 US-08-854-531-2 Sequence 2, Appl1
43 144 95.4 154 5 PCT-US95-13552-2 Sequence 2, Appl1
44 144 95.4 190 1 US-07-681-701-16 Sequence 16, Appl1
45 144 95.4 190 3 US-08-078-271B-1 Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-921-887-23
; Sequence 23, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAY
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-23

Query Match 100.0%; Score 151; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.3e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTIRRPQDVKFPGGQIVG 28
|||

Db 1 PKPQKTKRNTIRRPQDVKFPGGQIVG 28
|||

RESULT 2

US-08-290-665A-187
; Sequence 187, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R. H. AND
; APPLICANT: PURCELL, R. H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORGANISM: *homosapiens*
INDIVIDUAL ISOLATE: HK10
US-08-290-665A-187

Query Match 100.0%; Score 151; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. NO. 2.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels

Qy 1 PKPORKTKRNTIRRPQDVKFPGGGQIVG 28
|||||
Db 5 PKPORKTKRNTIRRPQDVKFPGGGQIVG 32
|||||

RESULT 3

US-08-290-665A-188
 ; Sequence 188, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R. H. AND
 ; APPLICANT: PURCELL, R. H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

```

, TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
, NUMBER OF SEQUENCES: 263
, CORRESPONDENCE ADDRESS:
, ADDRESSEE: MORGAN & FINNEGAN
, STREET: 345 PARK AVENUE
, CITY: NEW YORK
, STATE: NEW YORK
, COUNTRY: USA
, ZIP: 10154
, COMPUTER READABLE FORM:
, MEDIUM TYPE: FLOPPY DISK
, COMPUTER: IBM PC COMPATIBLE
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: WORDPERFECT 5.1
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/290,665A
, FILING DATE: 15-AUG-1994
, CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
US-08-290-665A-188

Query Match 100.0%; Score 151; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. NO. 2.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels

Qy 1 PKPQRKTKRNTIRRPQDVKFPGGQIVG 28
|||||

Db 5 PKPQRKTKRNTIRRPQDVKFPGGQIVG 32
|||||

RESULT 4
US-08-290-665A-190
; Sequence 190, Application US/08290665A
; Patent No. 5882852

GENERAL INFORMATION: BUKH, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF INVENTION: 263
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A

;; FILING DATE: 15-AUG-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK
;; REGISTRATION NUMBER: 36,459
;; REFERENCE/DOCKET NUMBER: 2026-4116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; TELEX: 421792
;; INFORMATION FOR SEQ ID NO: 190:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 191 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: DK12
;; -08-290-665A-190

Query Match 100.0%; Score 151; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFGGGQIVG 32

RESULT 5

PCT-US95-10398-187
; Sequence 187, Application PC/TUS9510398

;; GENERAL INFORMATION:
;; APPLICANT: BUKH, J., MILLER, R.H. AND
;; APPLICANT: PURCELL, R.H.
;; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
;; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
;; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
;; NUMBER OF SEQUENCES: 263
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORGAN & FINNEGAN
;; STREET: 345 PARK AVENUE
;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: USA
;; ZIP: 10154

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK
;; COMPUTER: IBM PC COMPATIBLE
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WORDPERFECT 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/10398
;; FILING DATE: 15-AUG-1995

;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/086,428

;; FILING DATE: 29 JUNE 1993

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/290/665

;; FILING DATE: 15 AUGUST 1994

;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK

;; REGISTRATION NUMBER: 36,459

;; REFERENCE/DOCKET NUMBER: 2026-4116

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800

;; TELEFAX: (212) 751-6849
;; TELEX: 421792

;; INFORMATION FOR SEQ ID NO: 187:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 191 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: HK10
;; PCT-US95-10398-187

Query Match 100.0%; Score 151; DB 5; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGGGQIVG 28
DB 5 PKPQRTKTRNTIRRPQDVKFGGGQIVG 32

RESULT 6

PCT-US95-10398-188

;; Sequence 188, Application PC/TUS9510398
;; GENERAL INFORMATION:

;; APPLICANT: BUKH, J., MILLER, R.H. AND
;; APPLICANT: PURCELL, R.H.

;; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

;; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

;; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

;; NUMBER OF SEQUENCES: 263

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORGAN & FINNEGAN

;; STREET: 345 PARK AVENUE
;; CITY: NEW YORK

;; STATE: NEW YORK
;; COUNTRY: USA

;; ZIP: 10154

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK

;; COMPUTER: IBM PC COMPATIBLE

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: WORDPERFECT 5.1

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/10398

;; FILING DATE: 15-AUG-1995

;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/086,428

;; FILING DATE: 29 JUNE 1993

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/290/665

;; FILING DATE: 15 AUGUST 1994

;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK

;; REGISTRATION NUMBER: 36,459

;; REFERENCE/DOCKET NUMBER: 2026-4116

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800

;; TELEFAX: (212) 751-6849
;; TELEX: 421792

;; INFORMATION FOR SEQ ID NO: 188:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 191 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: S52
;; PCT-US95-10398-188

Query Match 100.0%; Score 151; DB 5; Length 191;

LENGTH: 319
TYPE: PRT
ORGANISM: hepatitis C virus
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (144)..(144)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (149)..(149)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (156)..(157)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (167)..(167)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (171)..(172)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (174)..(174)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (177)..(177)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (232)..(232)
OTHER INFORMATION: Xaa is any amino acid
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (233)..(233)
OTHER INFORMATION: Xaa is any amino acid
US-08-635-886C-230

Query Match 100.0%; Score 151; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 4.5e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 PKPQKTKRNTIRPDVKFPGGGQIVG 28
|||||
5 PKPQKTKRNTIRPDVKFPGGGQIVG 32

RESULT 10
US-08-290-665A-189
Sequence 189, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA

ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-189

Query Match 99.3%; Score 150; DB 2; Length 191;
Best Local Similarity 96.4%; Pred. No. 3.6e-14;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRPDVKFPGGGQIVG 28
|||||
Db 5 PKPQKTKRNTIRPDVKFPGGGQIVG 32

RESULT 11
PCT-US95-10398-189
Sequence 189, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R. H. AND
APPLICANT: PURCELL, R. H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:

```
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
PCT-US95-10398-189

Query Match 99.3%; Score 150; DB 5; Length 191;
Best Local Similarity 96.4%; Pred. No. 3.6e-14;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGOIVG 28
Db 5 PKPQKTKRNTIRRPQDIKFGGGOIVG 32

RESULT 12
US-08-921-887-25
; Sequence 25, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEN, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis virus
US-08-921-887-25
```

```
Query Match 96.7%; Score 146; DB 3; Length 28;
Best Local Similarity 96.4%; Pred. No. 1.7e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGOIVG 28
Db 1 PKPQKTKRNTYRRPQDVKFGGGOIVG 28

RESULT 13
US-08-635-886C-191
; Sequence 191, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-191
```

```
Query Match 96.7%; Score 146; DB 4; Length 450;
Best Local Similarity 96.4%; Pred. No. 3.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGOIVG 28
Db 5 PKPQKTKRNTYRRPQDVKFGGGOIVG 32
```

```
RESULT 14
US-08-635-886C-192
; Sequence 192, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 192
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-192
```

```
Query Match 96.7%; Score 146; DB 4; Length 450;
Best Local Similarity 96.4%; Pred. No. 3.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 PKPQRTKRTIRRPQDVKFGGQIVG 28
|||||
Db 5 PKPQRTKRTIRRPQDVKFGGQIVG 32

RESULT 15
US-08-290-665A-175
: Sequence 175, Application US/08290665A
: Patent No. 5882852
: GENERAL INFORMATION:
: APPLICANT: BUKH, J., MILLER, R.H. AND
: APPLICANT: PURCELL, R.H.
: TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
: TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
: TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
: TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
: TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
: NUMBER OF SEQUENCES: 263
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/290,665A
: FILING DATE: 15-AUG-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: RICHARD W. BORK
: REGISTRATION NUMBER: 36,459
: REFERENCE/DOCKET NUMBER: 2026-4116
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: TELEX: 421792
: INFORMATION FOR SEQ ID NO: 175:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 191 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: ORIGINAL SOURCE:
: ORGANISM: homosapiens
: INDIVIDUAL ISOLATE: P8

US-08-290-665A-175
Query Match 96.0%; Score 145; DB 2; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.9e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTIRRPQDVKFGGQIVG 28
|||||
Db 5 PKPQRTKRTIRRPQDVKFGGQIVG 32

Search completed: August 7, 2003, 11:23:50
Job time : 11.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3635 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146a-23

Perfect score: 151

Sequence: 1 PPKQKTRNTRRRQDVKFGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues 451899

al number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	100.0	319	10	US-09-851-138-44
2	144	95.4	44	15	US-10-367-677-1
3	144	95.4	74	10	US-09-851-138-10
4	144	95.4	91	9	US-09-758-308-1
5	144	95.4	97	10	US-09-756-875-8
6	144	95.4	103	10	US-09-921-397-77
7	144	95.4	108	10	US-09-851-138-14
8	144	95.4	113	10	US-09-921-397-78
9	144	95.4	137	10	US-09-851-138-46
10	144	95.4	138	10	US-09-851-138-60
11	144	95.4	166	11	US-09-899-046-152
12	144	95.4	166	11	US-09-878-281-152
13	144	95.4	169	11	US-09-899-046-42
14	144	95.4	169	11	US-09-899-046-44
15	144	95.4	169	11	US-09-878-281-42

16 144 95.4 169 11 US-09-878-281-44
17 144 95.4 182 10 US-09-929-955-2
18 144 95.4 182 14 US-10-104-966-2
19 144 95.4 191 11 US-09-194-949-3
20 144 95.4 318 10 US-09-851-138-76
21 144 95.4 319 10 US-09-851-138-12
22 144 95.4 319 10 US-09-851-138-18
23 144 95.4 319 11 US-09-899-046-50
24 144 95.4 319 11 US-09-899-046-52
25 144 95.4 319 11 US-09-899-046-54
26 144 95.4 319 11 US-09-899-046-144
27 144 95.4 319 11 US-09-878-281-50
28 144 95.4 319 11 US-09-878-281-52
29 144 95.4 319 11 US-09-878-281-54
30 144 95.4 319 11 US-09-878-281-144
31 144 95.4 809 10 US-09-973-025-50
32 144 95.4 809 11 US-09-899-303-50
33 144 95.4 809 11 US-09-995-808-50
34 144 95.4 809 11 US-09-995-860-50
35 144 95.4 2894 10 US-09-941-611-23
36 144 95.4 2894 15 US-10-044-995-23
37 144 95.4 2985 15 US-10-259-275-40
38 144 95.4 3011 9 US-09-742-659-4
39 144 95.4 3011 10 US-09-952-572-9
40 144 95.4 3011 10 US-09-929-955-1
41 144 95.4 3011 10 US-09-747-419-20
42 144 95.4 3011 11 US-09-891-894-3
43 144 95.4 3011 14 US-10-104-966-1
44 144 95.4 3011 15 US-10-259-275-20
45 144 95.4 3012 10 US-09-238-076-2

ALIGNMENTS

RESULT 1

US-09-851-138-44
; Sequence 44, Application US/09851138
; Publication NO. US20020183508A1
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN

AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESS: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433

CITY: HOUSTON

STATE: TEXAS

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,138

FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>

APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994

APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 44:

```

; SEQUENCE CHARACTERISTICS:
;   LENGTH: 319 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-851-138-44

Query Match      100.0%; Score 151; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

RESULT 2
US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10/367,677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; TITLE: Of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match      95.4%; Score 144; DB 15; Length 44;
Best Local Similarity 96.4%; Pred. No. 1.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
Db 4 PKPQRTKTRNTIRRPQDVKPPGGQIVG 31

RESULT 3
US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US2002018308A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS

```

```

; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 74 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-851-138-10

Query Match      95.4%; Score 144; DB 10; Length 74;
Best Local Similarity 96.4%; Pred. No. 2.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

RESULT 4
US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US20020090607A1
; GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 14114.034902
; CURRENT APPLICATION NUMBER: US/09/758,308
; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match      95.4%; Score 144; DB 9; Length 91;
Best Local Similarity 96.4%; Pred. No. 2.8e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

```

RESULT 5

US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; STREET: Suite 701-E Columbia Square
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/756.875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 95.4%; Score 144; DB 10; Length 97;
Best Local Similarity 96.4%; Pred. No. 3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 PKPQKTKRNTIRRPQDVKFGGGQIVG 28
||||| ||||| ||||| ||||| ||||| |||||
5 PKPQKTKRNTIRRPQDVKFGGGQIVG 32

RESULT 6

US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-921-397-77

Query Match 95.4%; Score 144; DB 10; Length 103;
Best Local Similarity 96.4%; Pred. No. 3.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTIRRPQDVKFGGGQIVG 28
||||| ||||| ||||| ||||| ||||| |||||
Db 18 PKPQKTKRNTIRRPQDVKFGGGQIVG 45

RESULT 7

US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P. O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-851-138-14

Query Match 95.4%; Score 144; DB 10; Length 108;
Best Local Similarity 96.4%; Pred. No. 3.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTIRRPQDVKFGGGQIVG 28
||||| ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTIRRPQDVKFGGGQIVG 32

RESULT 8

US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS

; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921,397
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-78

Query Match 95.4%; Score 144; DB 10; Length 113;
Best Local Similarity 96.4%; Pred. No. 3.5e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
|||||
5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 9

US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:

Query Match 95.4%; Score 144; DB 10; Length 137;
Best Local Similarity 96.4%; Pred. No. 4.3e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
|||||
Db 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32
RESULT 10
US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO:
; AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 60:

Query Match 95.4%; Score 144; DB 10; Length 138;
Best Local Similarity 96.4%; Pred. No. 4.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKPPGGQIVG 28
|||||
Db 5 PKPQKTKRNTIRRPQDVKPPGGQIVG 32

RESULT 11

US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:

; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-152

```

```

Query Match          95.4%; Score 144; DB 11; Length 166;
Best Local Similarity 96.4%; Pred. No. 5.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

; 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
; ||||| ||||| ||||| ||||| |||||
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

```

```

RESULT 12
US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-152

```

```

Query Match          95.4%; Score 144; DB 11; Length 166;
Best Local Similarity 96.4%; Pred. No. 5.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

; 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
; ||||| ||||| ||||| ||||| |||||
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

```

```

RESULT 13
US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-42

```

```

Query Match          95.4%; Score 144; DB 11; Length 169;
Best Local Similarity 96.4%; Pred. No. 5.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

; 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
; ||||| ||||| ||||| ||||| |||||
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

```

```

RESULT 14
US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-44

```

```

Query Match          95.4%; Score 144; DB 11; Length 169;
Best Local Similarity 96.4%; Pred. No. 5.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

; 1 PKPQRTKTRNTIRRPQDVKPPGGQIVG 28
; ||||| ||||| ||||| ||||| |||||
Db 5 PKPQRTKTRNTIRRPQDVKPPGGQIVG 32

```

```

RESULT 15
US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270

```

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/878,281
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/362,455
;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 42:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 169 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-878-281-42

Query Match 95.4%; Score 144; DB 11; Length 169;
est Local Similarity 96.4%; Pred. No. 5.4e-13;
.atches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRPQDVKFPGGQIVG 28
DB 5 PKPQKTKRNTIRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 12:01:12
Job time : 15.3636 secs

RESULT 5

US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 97.4%; Score 149; DB 10; Length 97;
Best Local Similarity 96.4%; Pred. No. 1.7e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
||||:|||||
5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 6

US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE REFERENCE: B4809A - JAZ
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-921-397-77

Query Match 97.4%; Score 149; DB 10; Length 103;
Best Local Similarity 96.4%; Pred. No. 1.8e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
||||:|||||
Db 18 PKPQRTKNTNRRPDVKFPGGGQIVG 45

RESULT 7

US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-851-138-14

Query Match 97.4%; Score 149; DB 10; Length 108;
Best Local Similarity 96.4%; Pred. No. 1.9e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
||||:|||||
Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 8

US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS

Query Match	97.4%	Score 149;	DB 10;	Length 137;
Best Local Similarity	96.4%	Pred. No. 2.5e-13;		

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-152

Query Match 97.4%; Score 149; DB 11; Length 166;
Best Local Similarity 96.4%; Pred. No. 3e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 12

US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-152

Query Match 97.4%; Score 149; DB 11; Length 166;
Best Local Similarity 96.4%; Pred. No. 3e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 13

US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-42

Query Match 97.4%; Score 149; DB 11; Length 169;
Best Local Similarity 96.4%; Pred. No. 3.1e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 14

US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-44

Query Match 97.4%; Score 149; DB 11; Length 169;
Best Local Similarity 96.4%; Pred. No. 3.1e-13;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKNTNRRPDVKFPGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 15

US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-42

Query Match          97.4%; Score 149; DB 11; Length 169;
  sst Local Similarity 96.4%; Pred. No. 3.1e-13;
  atches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 PKPQRTKRNTRRPQDVKPPGGQIVG 28
      |||||:|||||:|||||:|||||:|||||
Db      5 PKPQRTKRNTRRPQDVKPPGGQIVG 32

Search completed: August 7, 2003, 12:01:12
Job time : 14.3636 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds

(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146A-33

Perfect score: 153

Sequence: 1 PRPQRTKRAHRRPDVRFPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

tal number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	153	100.0	28	AA1980	NC mosaic protein
2	148	96.7	28	AA1981	NC mosaic protein
3	141	92.2	312	AA1982	Hepatitis C virus
4	138	90.2	28	AA1983	NC mosaic protein
5	138	90.2	3010	AA1984	Blood transmissibl
6	138	90.2	3010	AA1985	HCV-S1 full-length
7	137	89.5	36	AA1986	CN14 fragment of H
8	137	89.5	38	AA1987	HCV capsid peptide
9	137	89.5	38	AA1988	HCV capsid peptide

```

10 137 89.5 38 14 AAR30689 HCV capsid peptide
11 137 89.5 38 15 AAR54065 Non-A, non-B hepat
12 137 89.5 43 19 AAW37380 Hepatitis C virus
13 137 89.5 44 19 AAW66083 Hepatitis C virus
14 137 89.5 44 20 AAY26952 Hepatitis C virus
15 137 89.5 44 21 AAY94410 Human hepatitis C
16 137 89.5 45 21 AAY94409 Human hepatitis C
17 137 89.5 55 13 AAR29534 HCV core-envelope
18 137 89.5 55 13 AAR29535 HCV core-envelope
19 137 89.5 55 13 AAR29536 HCV core-envelope
20 137 89.5 55 13 AAR29537 HCV core-envelope
21 137 89.5 55 13 AAR29542 HCV core-envelope
22 137 89.5 57 13 AAR25121 Non-A, Non-B Hepat
23 137 89.5 61 13 AAR20770 Peptide VIIIE base
24 137 89.5 61 16 AAR69545 Anti-HCV antibody
25 137 89.5 61 17 AAY15425 Prototype peptide
26 137 89.5 61 18 AAW01865 HCV core protein p
27 137 89.5 66 12 AAR12597 PT-NANB viral stru
28 137 89.5 74 17 AAR96530 Hepatitis C virus
29 137 89.5 78 13 AAR29160 Antigen pHCb101.
30 137 89.5 78 13 AAR29161 Antigen pHCb101.
31 137 89.5 79 14 AAR41422 HCV fragment 1 / I
32 137 89.5 79 14 AAR41423 HCV fragment 2 / I
33 137 89.5 80 15 AAR51389 Branched peptide H
34 137 89.5 90 16 AAR66619 Hepatitis C virus
35 137 89.5 91 22 AAB31695 Antigenic epitope
36 137 89.5 97 14 AAR40978 HCV core protein N
37 137 89.5 97 16 AAR66633 Hepatitis C virus
38 137 89.5 97 20 AAY01624 Protein encoded by
39 137 89.5 97 20 AAW30597 Hepatitis C virus
40 137 89.5 103 23 ABB77253 HCV bait polypepti
41 137 89.5 105 21 AAB18537 Protein encoded by
42 137 89.5 108 17 AAR96532 Hepatitis C virus
43 137 89.5 113 23 ABB77254 HCV bait polypepti
44 137 89.5 114 17 AAR98348 HCV capsid core pr
45 137 89.5 115 13 AAR29530 HCV core-envelope

```

ALIGNMENTS

RESULT 1

AA1980683
ID AAY06683 standard; Protein; 28 AA.

AC AAY06683;

DT 17-JUN-1999 (first entry)

DE NC mosaic protein amino acid fragment K.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
XX restriction endonuclease assisted ligation; vaccination.

OS Hepatitis C virus.

PN WO9910506-A1.

PD 04-MAR-1999.

PF 21-AUG-1998; 98WO-US17385.

PR 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1999-204671/17.

PT New mosaic protein, comprising a plurality of homologous antigenic
PT peptides from different genotypes of a species - useful for
PT detecting hepatitis infection in an individual

XX

PS Claim 5; Fig 9; 66pp; English.

XX

CC The invention relates to a mosaic protein, comprising a plurality of
 CC homologous antigenic peptides from different genotypes of a species. The
 CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
 CC synthesizing an artificial gene that encodes the mosaic protein is also
 CC provided. The method is designated restriction endonuclease assisted
 CC ligation (REAL). The mosaic protein and the artificial mosaic protein
 CC are useful for detecting a hepatitis infection in an individual. The
 CC mosaic gene and protein is also useful for vaccination against the
 CC infection, especially hepatitis C. The method of synthesizing the
 CC artificial gene and the resulting mosaic protein improve the sensitivity,
 CC spectrum of immunoreactivity, and antigen specificity of enzyme
 CC immunoassays. This provides improved detection of hepatitis C virus.
 CC Sequences AAY06673-683 represent amino acid sequence of each monomer
 CC comprising the NC mosaic protein.

XX

SQ Sequence 28 AA;

Query Match 100.0%; Score 153; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 2.6e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTRKKAHRRPDQVKFPGGGQIVG 28

||||| :||||| :||||| :||||| :|||||

Db 1 PKPQKTRKKAHRRPDQVKFPGGGQIVG 28

RESULT 2

AAY06678

ID AAY06678 standard; Protein; 28 AA.

XX

AC AAY06678;

XX

DT 17-JUN-1999 (first entry)

XX

DE NC mosaic protein amino acid fragment F.

XX

KW Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
 KW restriction endonuclease assisted ligation; vaccination.

XX

OS Hepatitis C virus.

XX

PN WO910506-A1.

XX

PD 04-MAR-1999.

XX

21-AUG-1998; 98WO-US17385.

XX

25-AUG-1997; 97US-0921887.

XX

(USSH) US DEPT HEALTH & HUMAN SERVICES.

XX

PI Fields HA, Khudiyakov YE;

XX

WPI; 1999-204671/17.

XX

PT New mosaic protein, comprising a plurality of homologous antigenic
 PT peptides from different genotypes of a species - useful for
 PT detecting hepatitis infection in an individual

XX

Claim 5; Fig 9; 66pp; English.

XX

CC The invention relates to a mosaic protein, comprising a plurality of
 CC homologous antigenic peptides from different genotypes of a species. The
 CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
 CC synthesizing an artificial gene that encodes the mosaic protein is also
 CC provided. The method is designated restriction endonuclease assisted
 CC ligation (REAL). The mosaic protein and the artificial mosaic protein
 CC are useful for detecting a hepatitis infection in an individual. The
 CC mosaic gene and protein is also useful for vaccination against
 CC infection, especially hepatitis C. The method of synthesizing the

CC

CC artificial gene and the resulting mosaic protein improve the sensitivity,
 CC spectrum of immunoreactivity, and antigen specificity of enzyme
 CC immunoassays. This provides improved detection of hepatitis C virus.
 CC Sequences AAY06673-683 represent amino acid sequence of each monomer
 CC comprising the NC mosaic protein.

XX SQ Sequence 28 AA;

Query Match 96.7%; Score 148; DB 20; Length 28;

Best Local Similarity 96.4%; Pred. No. 1.3e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRKKAHRRPDQVKFPGGGQIVG 28

||||| :||||| :||||| :||||| :|||||

Db 1 PKPQKTRKKAHRRPDQVKFPGGGQIVG 28

RESULT 3

AAR58593

ID AAR58593 standard; Protein; 312 AA.

XX

AC AAR58593;

XX

DT 09-MAY-1995 (first entry)

XX

DE Hepatitis C virus 349-1285 fragment antigen.

XX

KW Non-A non-B hepatitis virus antigens; NANBH; hepatitis C virus.

XX

OS Hepatitis C virus.

XX

PN JP06225770-A.

XX

PD 16-AUG-1994.

XX

PF 08-JUL-1993; 93JP-0193104.

XX

PR 10-JUL-1992; 92JP-0207391.

XX

(KOKU-) KOKUSAI SHIYAKU KK.

XX

PA (SANA) SANWA KAGAKU KENKYUSHO CO.

XX

PA (TOFU) TONEN CORP.

XX

PA (TOAR-) ZH TOKYO RINSHO IGAKU SGO KENKYUSHO.

XX

WPI; 1994-298800/37.

XX

N-PSDB; AAQ70543.

XX

A nucleic acid fragment coding Non-A Non-B Hepatitis virus

XX

antigens - for diagnosis of NANBH and detection of HCV

XX

Claim 1; Page 18; 22pp; Japanese.

XX

AAQ70543 is a fragment of hepatitis C virus (HCV) or non-A non-B

XX

hepatitis virus (NANBH) core and envelope one structural regions,

XX

encompassing base pairs 349-1285. It codes for AAR58593 an antigen

XX

to a structural region of the HCV virus, which can be used in

XX

the diagnosis of NANBH patients and the detection of HCV carriers.

XX

SQ Sequence 312 AA;

XX

Query Match 92.2%; Score 141; DB 15; Length 312;

XX

Best Local Similarity 92.9%; Pred. No. 1.5e-11;

XX

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

XX

QY 1 PKPQKTRKKAHRRPDQVKFPGGGQIVG 28

XX

Db 2 PKPQKTRKKAHRRPDQVKFPGGGQIVG 29

XX

RESULT 4

AAY06675

ID AAY06675 standard; Protein; 28 AA.

XX

PT Novel nucleic acid construct useful for detecting the presence of RNA virus, comprises an expression cassette and a promoter operably linked to expression cassette for minus strand RNA transcription of the cassette

XX

XX Example 1; Page 70-81; 81pp; English.

XX

CC The invention relates to nucleic acid construct which comprises an expression cassette including a first polynucleotide region including a 5' non-coding region (NCR) sequence of an RNA virus and at least an N-terminal portion of a coding sequence of RNA virus, a second polynucleotide region including a 3' untranslated region (UTR) sequence of the RNA virus and at least a C-terminal portion of a coding sequence of the virus and a third polynucleotide region encoding a reporter molecule, flanked by first and second polynucleotide regions; and a promoter sequence being operatively linked to expression cassette in a manner so as to enable a transcription of a minus strand RNA molecule from the expression cassette. Nucleic acid construct of the invention is useful for detecting the presence of an RNA virus in a cell. It is also useful for screening anti-viral drugs and determining drug resistance of an RNA virus. The present sequence is Hepatitis C virus (HCV) isolate HCV-S1 full-length polyprotein.

CC

XX Sequence 3010 AA;

QY Query Match 90.2%; Score 138; DB 23; Length 3010;
Best Local Similarity 89.3%; Pred. No. 4.1e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps

QY 1 PRPQRKTRKAHRRPDQVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PRPQRKTRNTYRRPDQVKFPGGGQIVG 32

RESULT 7

ID AAW06487

XX AAW06487 standard; Protein; 36 AA.

XX AAW06487;

XX

DT 31-JAN-1997 (first entry)

XX

DE CN14 fragment of HCV core region.

XX

XX CN14; CP14; core region; hepatitis C virus; HCV; detection; antibody.

KW

KW Synthetic.

OS

PN JP06327482-A.

XX

PD 29-NOV-1994.

XX

XX 21-MAY-1993; 93JP-0156026.

PF

XX

PR 21-MAY-1993; 93JP-0156026.

XX

XX (IMMO) IMMUNO JAPAN KK.

PA

XX WPI; 1995-047903/07.

DR

DR N-PSDB; AAT45055.

XX

XX Detection of hepatitis C virus - using oligopeptide fragment of HCV core region

PT

XX Claim 1; Page 6; 7pp; Japanese.

XX

XX This sequence is encoded by the oligonucleotide, CN14, and represents the peptide fragment CP14. CP14 is a fragment of the core region of hepatitis C virus (HCV). CP14 may be used in the detection of HCV infection and to raise antibodies against it.

CC

XX Sequence 36 AA;

SQ

Query Match 89.5%; Score 137; DB 16; Length 36;

Best Local Similarity 89.3%; Pred. No. 5.8e-12; Indels 0; Gaps 0;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVPPGGQIVG 28

Db 1 PKPQRTKRTNRRPDQVPPGGQIVG 28

RESULT 8

AAR30687

ID AAR30687 standard; peptide; 38 AA.

AC AAR30687;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 23.

KW Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS WO9222571-A1.

XX 23-DEC-1992.

XX 29-APR-1992; 92WO-US03635.

XX 13-JUN-1991; 91US-0714471.

PR 20-JUN-1991; 91US-0718052.

XX (BAXT) BAXTER DIAGNOSTICS INC.

XX Jolley ME, Leahy DC, Todd JA;

XX WPI; 1993-018073/02.

XX Synthetic peptide epitope with sequence encoded by hepatitis-C virus - for immunoassay for antigens for diagnosis of non-A, non-B hepatitis

XX Disclosure; Fig 1F; 66pp; English.

XX The sequences given in AAR30665-89 represent fragments of the Hepatitis C virus (HCV) amino acid sequence. They represent the beginning of the HCV open reading frame to amino acid 38 and encompass the "common" sequence. These peptides are contained in the capsid protein of the virus and themselves contain epitope groups. These peptides can be used in immunoassays for HCV antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and in competitive inhibition assay for detecting HCV specific antibodies.

CC (Updated on 25-MAR-2003 to correct PN field.)

XX Query Match 89.5%; Score 137; DB 14; Length 38;

Best Local Similarity 89.3%; Pred. No. 6.1e-12;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVPPGGQIVG 28

Db 5 PKPQRTKRTNRRPDQVPPGGQIVG 32

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30688 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30688 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30688 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30688 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30688 standard; peptide; 38 AA.

XX AAR30688;

AC 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 24.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS WO9222571-A1.

XX 23-DEC-1992.

XX 29-APR-1992; 92WO-US03635.

XX 13-JUN-1991; 91US-0714471.

PR 20-JUN-1991; 91US-0718052.

XX (BAXT) BAXTER DIAGNOSTICS INC.

XX Jolley ME, Leahy DC, Todd JA;

XX WPI; 1993-018073/02.

XX Synthetic peptide epitope with sequence encoded by hepatitis-C virus - for immunoassay for antigens for diagnosis of non-A, non-B hepatitis

XX Disclosure; Fig 1F; 66pp; English.

XX The sequences given in AAR30665-89 represent fragments of the Hepatitis C virus (HCV) amino acid sequence. They represent the beginning of the HCV open reading frame to amino acid 38 and encompass the "common" sequence. These peptides are contained in the capsid protein of the virus and themselves contain epitope groups. These peptides can be used in immunoassays for HCV antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and in competitive inhibition assay for detecting HCV specific antibodies.

CC (Updated on 25-MAR-2003 to correct PN field.)

XX Query Match 89.5%; Score 137; DB 14; Length 38;

Best Local Similarity 89.3%; Pred. No. 6.1e-12;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVPPGGQIVG 28

Db 5 PKPQRTKRTNRRPDQVPPGGQIVG 32

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

XX AAR30689;

XX 25-MAR-2003 (updated)

DT 11-MAY-1993 (first entry)

XX HCV capsid peptide No. 25.

XX Hepatitis C virus; HCV; open reading frame; "common" sequence; capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH; non-A, non-B hepatitis; competitive; inhibition assay.

XX Hepatitis C virus.

OS AAR30689 standard; peptide; 38 AA.

```

XX PN W09222571-A1.
XX PD 23-DEC-1992.
XX PF 29-APR-1992; 92WO-US03635.
XX PR 13-JUN-1991; 91US-0714471.
XX PR 20-JUN-1991; 91US-0718052.
XX PA (BAXT ) BAXTER DIAGNOSTICS INC.
XX PI Jolley ME, Leahy DC, Todd JA;
XX DR WPI; 1993-018073/02.
XX PT Synthetic peptide epitope with sequence encoded by hepatitis-C
PT virus - for immunoassay for antigens for diagnosis of non-A,
PT non-B hepatitis
XX Disclosure; Fig 1F; 66pp; English.

The sequences given in AAR30665-89 represent fragments of the
CC Hepatitis C virus (HCV) amino acid sequence. They represent the
CC beginning of the HCV open reading frame to amino acid 38 and
CC encompass the "common" sequence. These peptides are contained in
CC the capsid protein of the virus and themselves contain epitope
CC groups. These peptides can be used in immunoassays for HCV
CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
CC in competitive inhibition assay for detecting HCV specific
CC antibodies.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 38 AA;
SQ
Query Match 89.5%; Score 137; DB 14; Length 38;
Best Local Similarity 89.3%; Pred. No. 6.1e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKAHRPPQDVKFPGGGQIVG 28
Db ||||||| :|||||||||
5 PKPQRTKTRNTNRRPQDVKFPGGGQIVG 32

RESULT 11
AAR54065
ID AAR54065 standard; Protein; 38 AA.
XX AAR54065;
14-FEB-1995 (first entry)
XX Non-A, non-B hepatitis virus coreI region fragment.
XX DE
XX Non-A, non-B hepatitis virus; NANBH; hepatitis C virus; HCV;
KW core; ENV; NS1; NS2; NS3; antigen; detection.
XX OS Hepatitis C virus.
XX JP06141870-A.
XX PD 24-MAY-1994..
XX PF 12-MAR-1992; 92JP-0088140.
XX PR 12-MAR-1992; 92JP-0088140.
XX PA (SANW ) SANWA KAGAKU KENKYUSHO CO.
XX PA (TOKR-) ZH TOKYO RINSHO IGAKU SOGO KENKYUSHO.
XX DR WPI; 1994-205028/25.
DR N-PSDB; AAQ64067.

```

```

XX DNA coding a Non-A, non-B hepatitis virus antigen - useful for
PT detecting HCV within serum
XX Disclosure; Page 11; 22pp; Japanese.
XX Hepatitis C virus #4 and #6 genes were isolated (AAQ64068-69).
CC Both genes contain the core, ENV, NS1, NS2 and NS3 regions.
CC A core region fragment is given in AAQ64067.
XX Sequence 38 AA;
SQ
Query Match 89.5%; Score 137; DB 15; Length 38;
Best Local Similarity 89.3%; Pred. No. 6.1e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKAHRPPQDVKFPGGGQIVG 28
Db ||||||| :|||||||||
5 PKPQRTKTRNTNRRPQDVKFPGGGQIVG 32

RESULT 12
AAW37380
ID AAW37380 standard; Protein; 43 AA.
XX AAW37380;
XX 11-MAR-1998 (first entry)
XX Hepatitis C virus C-1 protein 1-43.
XX Hepatitis C virus; HCV; chimeric; antigen; detection; core region;
KW epitope; NS3; NS4; infection.
XX OS Hepatitis c virus.
XX JP09278794-A.
XX PD 28-OCT-1997.
XX PF 10-FEB-1997; 97JP-0027015.
XX PR 09-FEB-1996; 96JP-0024045.
XX PA (TOFU ) TONEN CORP.
XX DR WPI; 1998-022248/03.
XX PT New chimeric peptide antigen derived from hepatitis C virus protein
PT - useful for detecting HCV infections
XX PS Disclosure; Page 24; 30pp; Japanese.
XX The present sequence represents a Hepatitis C virus (HCV) protein
CC sequence from the disclosure of the present specification. The
CC present specification describes a chimeric HCV peptide antigen which
CC comprises at least 2 peptide epitope regions from the HCV polypeptide
CC core region, 2 peptide epitope regions from the NS3 region and at
CC least 2 peptide epitope regions from the NS4 region. The antigen binds
CC specifically with an antibody produced by a human infected by HCV. The
CC peptide can detect a wide range of HCV infections with high sensitivity.
XX Sequence 43 AA;
SQ
Query Match 89.5%; Score 137; DB 19; Length 43;
Best Local Similarity 89.3%; Pred. No. 7e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKAHRPPQDVKFPGGGQIVG 28
Db ||||||| :|||||||||
5 PKPQRTKTRNTNRRPQDVKFPGGGQIVG 32

```

```

RESULT 13
AAW66083
ID AAW66083 standard; peptide: 44 AA.
XX
AC AAW66083;
XX
DT 16-NOV-1998 (first entry)
XX
DE Hepatitis C virus p21 protein N-terminal fragment.
XX
KW antigenic; nucleocapsid; p21 protein; alpha-helical; immunogen;
KW antibody; hepatitis C virus; epitope; N-terminal fragment.
XX
OS hepatitis c virus.
XX
PN WO9839360-AL.
XX
PD 11-SEP-1998.
XX
05-MAR-1998; 98WO-FR00442.
05-MAR-1997; 97FR-0002878.
(INMR ) BIO MERIEUX.
XX
PA Dalbon P, Jolivet M, Lacoux X, Ladaviere L, Penin F;
XX
PI WPI; 1998-495793/42.
XX
DR
XX
PT New peptide from the N-terminus of hepatitis C virus p21 protein
PT containing the immunodominant epitope - and related antibodies, used
PT for diagnosis, treatment and prevention of hepatitis C infection
XX
XX
PS Disclosure; Page 16; 37pp; French.
XX
XX
CC The invention relates to a peptide, which is recognised by antibodies
CC against amino acids 2-45 at the N-terminus of the core (or nucleocapsid)
CC p21 protein of hepatitis C virus (HCV), or its variants. The peptide has
CC a tertiary structure consisting of two alpha-helical fragments, almost
CC perpendicular to each other in space, connected by a junction peptide.
CC Excluded are all proteins and peptides comprising, or consisting of, the
CC N-terminal part of p21 (starting from amino acid 1 or 2). Also new are
CC (1) monoclonal or polyclonal antibodies produced using the peptide as
CC an immunogen and (2) complex consisting of the peptide specifically
CC bound to some other molecule, particularly peptide or nucleotide
CC fragments or functionalised aromatic compounds. The peptide is used
CC as immunogen for generating antibodies and (ii) for detecting and
CC quantifying either antibodies against p21 or HCV-derived mRNA (by
CC complex formation). Antibodies are used correspondingly to detect HCV or
CC related antigens. The peptide and antibodies may also be used to treat
CC or prevent HCV infections. The present sequence represents the N-terminus
CC of the core p21 protein of hepatitis C virus.
XX
SQ Sequence 44 AA;
XX
Query Match 89.5%; Score 137; DB 19; Length 44;
Best Local Similarity 89.3%; Pred. No. 7.1e-12; Indels 0; Gaps 0;
Matches 25; Conservative 1; Mismatches 2;
Oy 1 PKPQKTKRKAHRRPDVKFPGGGQIVG 28
Db 4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31
RESULT 14
AAW26952
ID AAW26952 standard; peptide: 44 AA.
XX
AC AAW26952;
XX
XX
DT 21-DEC-1999 (first entry)
XX
DE Hepatitis C virus Core protein amino acids 2-45.

```

```

XX
KW Epitope; hepatitis C virus; core protein; monoclonal antibody;
KW diagnosis; infection; sandwich immunoassay.
XX
OS Hepatitis C virus.
XX
PN FR2775690-AL.
XX
PD 10-SEP-1999.
XX
PF 09-MAR-1998; 98FR-0003087.
XX
PR 09-MAR-1998; 98FR-0003087.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Jolivet RC, Piga N, Yvon S, Paranhos BC, Jolivet M;
XX
DR WPI; 1999-530397/45.
XX
PT Monoclonal antibodies useful for detecting and/or quantifying hepatitis
PT C virus core protein
XX
PS Claim 3; Page 10; 19pp; French.
XX
CC Peptides AAY26949-Y26955 represent peptide epitopes derived from the
CC N-terminus of the hepatitis C virus core protein. The peptides are used
CC to generate monoclonal antibodies or antibody fragments specific for
CC hepatitis C virus (HCV) core protein. The monoclonal antibodies are
CC used for early diagnosis of HCV infections, especially by two-antibody
CC sandwich immunoassay.
XX
SQ Sequence 44 AA;
XX
Query Match 89.5%; Score 137; DB 20; Length 44;
Best Local Similarity 89.3%; Pred. No. 7.1e-12; Indels 0; Gaps 0;
Matches 25; Conservative 1; Mismatches 2;
Oy 1 PKPQKTKRKAHRRPDVKFPGGGQIVG 28
Db 4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31
RESULT 15
AAW94410
ID AAY94410 standard; peptide: 44 AA.
XX
AC AAY94410;
XX
DT 11-SEP-2000 (first entry)
XX
DE Human hepatitis C virus core protein N-terminus, residues 2-45.
XX
KW Human; hepatitis C virus; HCV; immunogen; antigen; antibody; virucide;
KW hepatotropic; anti-inflammatory; virus detection; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200031130-A1.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-IB01933.
XX
PR 20-NOV-1998; 98US-0196155.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Dalbon P, I Dalbon P, Jolivet M, Jolivet-Reynaud C;
XX
DR WPI; 2000-411934/35.
XX
PT Polypeptides that bind to anti-hepatitis C virus antibodies, useful for

```

PT diagnosing and preventing hepatitis C infections -
XX
PS Claim 1: Page 42; 50pp; English.
XX
CC The present peptide, designated S42G, corresponds to residues 2 to 45 of
CC the N-terminus of the human hepatitis C virus (HCV) core protein. It is
CC an immunodominant region containing conformational type epitopes and
CC linear type epitopes. It manifests an immunoreactivity with all the sera
CC of individuals or blood samples infected with HCV and which possess
CC antibodies directed against the core protein. An amino acid may be
CC substituted for homologous amino acids and side chains and peptide bonds
CC may be modified. For example, L-amino acids may be replaced by D-amino
CC acids, amine groups may be acetylated, and so on. The native antigenic
CC sequence and its antigenic derivatives may be used for detection of
CC hepatitis C virus and for raising antibodies against the virus.
XX
SQ Sequence 44 AA;

Query Match 89.5%; Score 137; DB 21; Length 44;
Best Local Similarity 89.3%; Pred. No. 7.1e-12;
atches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

W 1 PKPQKTRKKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQKTRKNTNRFPQDKFPGGGQIVG 31

Search completed: August 7, 2003, 11:14:10
Job time : 38.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-33
Perfect score: 153
Sequence: 1 PKPQRTKRAHRRPQDVKPPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	90.2	3010	1 A45573	genome polyprotein
2	137	89.5	108	2 S41353	genome polyprotein
3	137	89.5	108	2 S41355	genome polyprotein
4	137	89.5	108	2 S41357	genome polyprotein
5	137	89.5	108	2 S41348	genome polyprotein
6	137	89.5	112	2 S41371	genome polyprotein
7	137	89.5	112	2 S41341	genome polyprotein
8	137	89.5	114	2 S41370	genome polyprotein
9	137	89.5	114	2 S41369	genome polyprotein
10	137	89.5	114	2 S41368	genome polyprotein
11	137	89.5	115	2 S41342	genome polyprotein
12	137	89.5	115	2 S41344	genome polyprotein
13	137	89.5	115	2 S41350	genome polyprotein
14	137	89.5	115	2 S41354	genome polyprotein
15	137	89.5	115	2 S41345	genome polyprotein
16	137	89.5	115	2 S41347	genome polyprotein
17	137	89.5	115	2 S41343	genome polyprotein
18	137	89.5	118	2 S41346	genome polyprotein
19	137	89.5	369	2 S21471	genome polyprotein
20	137	89.5	441	2 S12707	genome polyprotein
21	137	89.5	513	2 PC1284	genome polyprotein
22	137	89.5	520	2 J01925	polyprotein - hepa
23	137	89.5	523	2 J01926	polyprotein - hepa
24	137	89.5	550	2 JH0711	genome polyprotein
25	137	89.5	782	2 S19876	genome polyprotein
26	137	89.5	782	2 S18031	genome polyprotein
27	137	89.5	782	2 S18032	genome polyprotein
28	137	89.5	787	2 P00677	hypothetical prote
29	137	89.5	874	2 J00883	genome polyprotein

30 137 89.5 874 2 J00881 genome polyprotein
31 137 89.5 876 2 PC2219 polypeptide - hepa
32 137 89.5 3010 1 GNVVTC genome polyprotein
33 137 89.5 3010 1 GNVVTC genome polyprotein
34 137 89.5 3010 1 S18030 genome polyprotein
35 137 89.5 3011 1 GNVVCH genome polyprotein
36 137 89.5 3011 1 S40770 genome polyprotein
37 137 89.5 3014 1 JC5620 genome polyprotein
38 137 89.5 3033 1 J01303 genome polyprotein
39 137 89.5 3033 1 GNVVJ8 genome polyprotein
40 135 88.2 88 2 S21336 genome polyprotein
41 133 86.9 108 2 S41356 genome polyprotein
42 133 86.9 114 2 S41359 genome polyprotein
43 133 86.9 114 2 S41358 genome polyprotein
44 133 86.9 115 2 S41351 genome polyprotein
45 133 86.9 115 2 S41349 genome polyprotein

ALIGNMENTS

RESULT 1
A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hiji
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carr
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; NID:G221612; PIDN:BAA01943.1; PID:G22161
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-101/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 90.2%; Score 138; DB 1; Length 3010;
Best Local Similarity 89.3%; Pred. No. 2.5e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKRAHRRPQDVKPPGGQIVG 28
DB 5 PKPQRTKRTYRRPQDVKPPGGQIVG 32

RESULT 2
S41353
genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41353
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341

A:Accession: S41353
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:229456
 A:Experimental source: genotype 2, N2
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 108;
 Best Local Similarity 89.3%; Pred. No. 1.6e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

JLT 3

S41355
 genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N4
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41355
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41355
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:229458
 A:Experimental source: genotype 2, N4
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 108;
 Best Local Similarity 89.3%; Pred. No. 1.6e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
 ||||| :|||||
 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 4

S41357
 genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N6
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41357
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41357
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:229460
 A:Experimental source: genotype 2, N6
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 108;

Best Local Similarity 89.3%; Pred. No. 1.6e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 5

S41348
 genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N6
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41348
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41348
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:229451
 A:Experimental source: genotype 1, N6
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 108;
 Best Local Similarity 89.3%; Pred. No. 1.6e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 6

S41371
 genome polyprotein - hepatitis C virus (genotype 5, N5) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 5, N5
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41371
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41371
 A:Molecule type: genomic RNA
 A:Residues: 1-112 <VAN>
 A:Cross-references: EMBL:229474
 A:Experimental source: genotype 5, N5
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-112/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 112;
 Best Local Similarity 89.3%; Pred. No. 1.6e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 7

S41341
 genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus

A:Variety: genotype 1, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41341
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
C:Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>
A:Accession: S41341
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
A:Cross-references: EMBL:229444; NID:g443850; PIDN:CAA82582.1; PID:g443851
A:Experimental source: genotypel, N1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 112;
Best Local Similarity 89.3%; Pred. No. 1.6e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 PKPQKTKRKAHRRPQDVKFFGGGQIVG 28
||||||| :|||||||
5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32

RESULT 8
S41370
genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41370
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41370
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229473; NID:g443908; PIDN:CAA82611.1; PID:g443909
A:Experimental source: genotype 5, N4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 1.6e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPQDVKFFGGGQIVG 28
||||||| :|||||||
Db 5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32

RESULT 9
S41369
genome polyprotein - hepatitis C virus (genotype 5, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41369
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41369
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229472; NID:g443906; PIDN:CAA82610.1; PID:g443907
A:Experimental source: genotype 5, N3
C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 1.6e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPQDVKFFGGGQIVG 28
||||||| :|||||||
Db 5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32

RESULT 10
S41368
genome polyprotein - hepatitis C virus (genotype 5, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41368
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41368
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229471; NID:g443904; PIDN:CAA82609.1; PID:g443905
A:Experimental source: genotype 5, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 1.6e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPQDVKFFGGGQIVG 28
||||||| :|||||||
Db 5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32

RESULT 11
S41342
genome polyprotein - hepatitis C virus (genotype 1, N10) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N10
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41342
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41342
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229445; NID:g443852; PIDN:CAA82583.1; PID:g443853
A:Experimental source: genotype 1, N10
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.5%; Score 137; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 1.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPQDVKFFGGGQIVG 28
||||||| :|||||||
Db 5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32

RESULT 12

```
S41344
genome polyprotein - hepatitis C virus (genotype 1, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41344
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41344
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229447; NID:g443856; PIDN:CAA82585.1; PID:g443857
A:Experimental source: genotype 1, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match      89.5%; Score 137; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 1.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 13
S41350
genome polyprotein - hepatitis C virus (genotype 1, N8) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N8
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41350
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41350
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229453; NID:g443868; PIDN:CAA82591.1; PID:g443869
A:Experimental source: genotype 1, N8
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match      89.5%; Score 137; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 1.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 14
S41354
genome polyprotein - hepatitis C virus (genotype 2, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41354
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41354
A:Molecule type: genomic RNA
```

```
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229457
A:Experimental source: genotype 2, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match      89.5%; Score 137; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 1.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 15
S41345
genome polyprotein - hepatitis C virus (genotype 1, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41345
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41345
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229448; NID:g443858; PIDN:CAA82586.1; PID:g443859
A:Experimental source: genotype 1, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match      89.5%; Score 137; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 1.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKTRNTNRRPDQVKFPGGGQIVG 32

Search completed: August 7, 2003, 11:21:49
Job time : 9.54545 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run On: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146A-33
Perfect score: 153
Sequence: 1 PKPQTKRKAHRRPDQVFKPGGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41: *
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	138	90.2	3010	1	POLG_HCVJT
2	137	89.5	513	1	POLG_HCVJ2
3	137	89.5	520	1	POLG_HCVH4
4	137	89.5	520	1	POLG_HCVHK
5	137	89.5	737	1	POLG_HCVJ5
6	137	89.5	737	1	POLG_HCVJ7
7	137	89.5	3010	1	POLG_HCVBK
8	137	89.5	3010	1	POLG_HCVJA
9	137	89.5	3011	1	POLG_HCVH
10	137	89.5	3033	1	POLG_HCVJ6
11	137	89.5	3033	1	POLG_HCVJ8
12	130	85.0	3010	1	POLG_HCVTW
13	129	84.3	3011	1	POLG_HCV1
14	53	34.6	512	1	IE63_HSV2H
15	53	34.6	794	1	FURI_HUMAN
16	51	33.3	282	1	RK4_TOBAC
17	51	33.3	454	1	NCAP_CVM3
18	51	33.3	454	1	NCAP_CVM5
19	51	33.3	454	1	NCAP_CVM5
20	50	32.7	293	1	RK4_SPIOL
21	48	31.4	153	1	RL29_MYCGA
22	48	31.4	793	1	FURI_MOUSE
23	47	30.7	68	1	ATP8_PONPP
24	47	30.7	68	1	ATP8_PONPY
25	47	30.7	530	1	F262_BOVIN
26	47	30.7	538	1	PIG2_YEAST
27	47	30.7	708	1	GBF_DICDI
28	47	30.7	797	1	FURI_BOVIN
29	47	30.7	1531	1	YQ38_CAEEL
30	46	30.1	139	1	KMLS_PIG
31	46	30.1	194	1	RS7_FUGRU
32	46	30.1	194	1	RS7_HUMAN
33	46	30.1	436	1	GDF6_BOVIN

34	46	30.1	449	1	APG_BRANA
35	46	30.1	455	1	NCAP_CVMJH
36	46	30.1	556	1	PDPK_HUMAN
37	46	30.1	559	1	PDPK_MOUSE
38	46	30.1	559	1	PDPK_RAT
39	46	30.1	857	1	KCB1_MOUSE
40	46	30.1	1176	1	KMLS_BOVIN
41	46	30.1	1211	1	PKP4_HUMAN
42	46	30.1	1616	1	SLAP_BACCI
43	46	30.1	1914	1	KMLS_HUMAN
44	45.5	29.7	740	1	Y179_HUMAN
45	45	29.4	189	1	RS7A_YEAST

ALIGNMENTS

RESULT 1
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RL infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D11168; BAA01943.1;
DR PIR; A45573; A45573.
DR PDB; 1A1Q; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR MEROPS; S29.001;
DR MEROPS; U39.001;
DR InterPro; IPR001410; DEAD.

```

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR001666; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_Psivir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR Core protein; RNA-directed RNA polymerase;
KW Core protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 384 729 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 730 1006 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 1616 1862 NON-STRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 1863 2013 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT CHAIN 2014 3010 POTENTIAL.
FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1165 1165 ATP (POTENTIAL).
FT SITE 1230 1237 DECH_BOX.
FT CARBOHYD 1316 1319 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MD; 94A1C7743D642BB CRC64;

Query Match 90.2%; Score 138; DB 1; Length 3010;
Best Local Similarity 89.3%; Pred. No. 2e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRKTRKARRPDQVKFPGGQIVG 28

```

```

Db 5 PKPQRKTRKARRPDQVKFPGGQIVG 32
RESULT 2
POLG_HCVJ2
ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed-1314459;
RA Okamoto H., Kural K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10074; BRA00968.1;
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 >513 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 513 513

```

```

FT  NON_TER      520
SQ  SEQUENCE      520 AA; 56499 MW; AAL35246CF20D525 CRC64;

Query Match      89.5%; Score 137; DB 1; Length 520;
Best Local Similarity 89.3%; Pred. No. 4.4e-12;
Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy  1 PKPQRTKRRKAHRRPDQVKFPGGGQIVG 28
    ||||| :|||||
Db   5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32

RESULT 4
POLG_HCVHK
ID  POLG_HCVHK      STANDARD;      PRT;      520 AA.
AC  O01403;
DT  01-JUL-1993 (Rel. 26, Created)
DT  01-JUL-1993 (Rel. 26, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE  Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE  (GP68) (GP70) (NS1); Fragment].
OS  Hepatitis C virus (isolate HCV-KF) (HCV).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
NCBI_TaxID=31644;
[1]
SEQUENCE FROM N.A.
MEDLINE=93019030; PubMed=1383400;
Abe K., Inchauspe G., Fujisawa K.;
"Genomic characterization and mutation rate of hepatitis C virus
isolated from a patient who contracted hepatitis during an epidemic
of non-A, non-B hepatitis in Japan.";
J. Gen. Virol. 73:2725-2729(1992).
CC  -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC  LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC  PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC  PROTEIN C AND MRNA.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; D10687; BAA01529.1; --
DR  PIR; JQ1925; JQ1925.
DR  InterPro; IPR002522; HCV_capsid.
DR  InterPro; IPR002521; HCV_Core.
DR  InterPro; IPR002519; HCV_env.
DR  InterPro; IPR002531; HCV_NSI.
DR  Pfam; PF01543; HCV_capsid; 1.
DR  Pfam; PF01542; HCV_core; 1.
DR  Pfam; PF01539; HCV_env; 1.
DR  Pfam; PF01560; HCV_NSI; 1.
DR  ProDom; PD186062; HCV_NSI; 1.
KW  Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW  Transmembrane; Nonstructural protein.
FT  INIT_MET      1 1
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  TRANSMEM      347 369
FT  CARBOHYD      196 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383

```

```

FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FE27349B CRC64;

Query Match
Best Local Similarity 89.5%; Score 137; DB 1; Length 520;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFGGQIIVG 28
    ||||| . . . |||||
DB 5 PKPQKTKRNTNRRPDQVKFGGQIIVG 32

RESULT 5
POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (Isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10075; BAA00969.1;
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF0186062; HCV_NS1; 1.
CC Pfam; PF0186062; HCV_NS1; 1.
CC Transmembrane; Glycoprotein; Coat protein; Envelope protein;
KW Polyprotein; Nonstructural
KW Transmembrane; Nonstructural
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737

```

```

FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match
Best Local Similarity 89.5%; Score 137; DB 1; Length 737;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFGGQIIVG 28
    ||||| . . . |||||
DB 5 PKPQKTKRNTNRRPDQVKFGGQIIVG 32

RESULT 6
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (Isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10077; BAA00971.1;
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC Pfam; PF01543; HCV_capsid; 1.

```

DR Pfam: PF01542; HCV_core: 1.
 DR Pfam: PF01539; HCV_env: 1.
 DR Pfam: PF01560; HCV_NS1: 1.
 DR ProDom: PD186062; HCV_NS1: 1.
 KW Transmembrane; Glycoprotein; Coat protein; Envelope protein;
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match
 Best Local Similarity 89.5%; Score 137; DB 1; Length 737;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRAHRRPQDVKPPGGQIVG 28
 DB 5 PKPQKTKRNTNRPPQDVKPPGGQIVG 32

RESULT 7
 POLG_HCVBK STANDARD; PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
 NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 NS5A (P56); Nonstructural protein NS5B (P56); Nonstructural protein
 NS5C (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 from human carriers".
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits

phosphorylation mediated by cAMP-dependent protein kinase.";
 Eur. J. Biochem. 237:611-618(1996).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RA MEDLINE=97015088; PubMed=8861916;
 RX Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moosaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 trypsin-like fold and a structural zinc binding site.";
 Cell 87:331-342(1996).
 [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RA MEDLINE=98227846; PubMed=9568891;
 RX Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of HCV strain hepatitis C
 virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the p6
 position. Cys or Thr in Pl and Ser or Ala in Pl'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS.
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: M58335; AAA72945.1;
 DR PIR: A38465; GNVVTC.
 DR PDB: 1A10; 25-MAR-98.
 DR PDB: 1JXP; 14-JAN-98.
 DR PDB: 1NS3; 08-APR-98.
 DR PDB: 1C2P; 15-NOV-00.
 DR PDB: 1CSJ; 08-NOV-99.
 DR PDB: 1GX5; 09-APR-02.
 DR PDB: 1GX6; 10-APR-02.
 DR PDB: 1QV; 26-JUN-00.
 DR PDB: 1QV; 26-JUN-00.
 DR PDB: 1QV; 26-JUN-00.
 DR MEROPS: S29.001;
 DR MEROPS: U39.001;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS-PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 FT STRAND 1031 1035
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186

Best Local Similarity 89.3%; Score 137; DB 1; Length 3010;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRAHRRPDQVKKFGGGQIVG 28
 DB 5 PKPQRTKRTNRRPDQVKKFGGGQIVG 32

RESULT 8
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraliso K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in Pl and Ser or Ala in Pl'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90208; BAA14233.1;
 DR PIR; A39253; GNVVCJ.

HSP: P26663; 1JXP.
 MEROPS: S29.001; -
 MEROPS: U39.001; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PS_vir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01540; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein: Glycoprotein; Transferase: RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural
 KW INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 369
 FT TRANSEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 SITE 1316 1319
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 250 250
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 532 532
 CARBOHYD 556 556
 CARBOHYD 576 576
 CARBOHYD 623 623
 CARBOHYD 645 645
 CARBOHYD 2041 2041
 CARBOHYD 2077 2077
 CARBOHYD 2240 2240
 CARBOHYD 2788 2788
 SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 89.5%; Score 137; DB 1; Length 3010;
 Best Local Similarity 89.3%; Pred. No. 2.8e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRAHRRPQDVKFPGGQIVG 28
 ||||| :|||||||
 Db 5 PKPQRTKRTNRRPQDVKFPGGQIVG 32

RESULT 9
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92052256; PubMed-1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE-9731322; PubMed-9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE-98154321; PubMed-9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 Murchio M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -|- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -|- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -|- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -|- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -|- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -|- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -|- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -|- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M67463; AAA4534.1; -

PIR; A36814; GNVVCH.

PDB; 1HEI; 25-NOV-98.

PDB; 1AIV; 16-FEB-99.

PDB; 1AIR; 17-JUN-98.

MEROPS; S29.001; -

MEROPS; U39.001; -

TRANSFAC; T04155; -

InterPro; IPR001410; DEAD.

InterPro; IPR002522; HCV_capsid.

InterPro; IPR002521; HCV_core.

InterPro; IPR002519; HCV_env.

InterPro; IPR002531; HCV_NSI.

InterPro; IPR002518; HCV_NS2.

InterPro; IPR004109; HCV_NS3.

InterPro; IPR000745; HCV_NS4a.

InterPro; IPR001490; HCV_NS4b.

InterPro; IPR002868; HCV_NS5a.

InterPro; IPR002166; HCV_RdRP.

InterPro; IPR001650; Helicase_C.

InterPro; IPR007095; RNA_pol_DS_PS.

InterPro; IPR007094; RNA_pol_PSVir.

Pfam; PF01543; HCV_capsid; 1.

Pfam; PF01542; HCV_core; 1.

Pfam; PF01539; HCV_env; 1.

Pfam; PF01560; HCV_NSI; 1.

Pfam; PF01538; HCV_NS2; 1.

Pfam; PF02907; HCV_NS3; 1.

Pfam; PF01006; HCV_NS4a; 1.

Pfam; PF01001; HCV_NS4b; 1.

Pfam; PF01506; HCV_NS5a; 1.

Pfam; PF00271; Helicase_C; 1.

Pfam; PF00998; Viral_RdRP; 1.

ProDom; PD186062; HCV_NSI; 1.

SMART; SM00487; DEXDC; 1.

Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

3D-structure.

INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

CELLULAR AMINOPEPTIDASE.

CHAIN 1 191 CAPSID PROTEIN C.

CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.

CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.

CHAIN 747 809 PROTEIN P7.

CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.

CHAIN 1027 1657 PROTEASE/HELICASE NS3.

CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.

CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.

CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.

CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.

CHAIN 3012 369 POTENTIAL.

CHAIN 370 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

CHAIN 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

CHAIN 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

CHAIN 1230 1237 ATP (POTENTIAL).

CHAIN 1316 1319 DECH_BOX.

CHAIN 1319 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

CHAIN 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	623	623
FT	CARBOHYD	645	645
FT	STRAND	1224	1226
FT	TURN	1232	1233
FT	TURN	1236	1238
FT	HELIX	1239	1246
FT	TURN	1247	1248
FT	STRAND	1251	1255
FT	HELIX	1258	1271
FT	TURN	1272	1272
FT	STRAND	1277	1280
FT	TURN	1281	1282
FT	STRAND	1283	1285
FT	STRAND	1291	1295
FT	HELIX	1296	1301
FT	TURN	1302	1303
FT	STRAND	1312	1316
FT	TURN	1317	1319
FT	HELIX	1323	1335
FT	TURN	1336	1340
FT	STRAND	1343	1347
FT	TURN	1352	1353
FT	TURN	1360	1361
FT	STRAND	1362	1366
FT	STRAND	1368	1368
FT	STRAND	1373	1375
FT	TURN	1376	1377
FT	STRAND	1378	1380
FT	HELIX	1382	1385
FT	STRAND	1389	1393
FT	HELIX	1397	1409
FT	TURN	1410	1411
FT	STRAND	1414	1417
FT	TURN	1419	1420
FT	STRAND	1432	1436
FT	TURN	1438	1439
FT	STRAND	1450	1453
FT	STRAND	1456	1463
FT	STRAND	1471	1478
FT	STRAND	1480	1480
FT	HELIX	1481	1488
FT	TURN	1489	1490
FT	STRAND	1497	1501
FT	STRAND	1507	1507
FT	STRAND	1511	1511
FT	HELIX	1514	1527
FT	STRAND	1532	1544
FT	STRAND	1550	1550
FT	HELIX	1555	1564
FT	HELIX	1570	1578
FT	TURN	1579	1580
FT	HELIX	1584	1597
FT	TURN	1598	1598
FT	HELIX	1606	1611
FT	TURN	1614	1618
FT	STRAND	1622	1623
FT	STRAND	1627	1627
FT	STRAND	1635	1636
FT	HELIX	1640	1652
SQ	SEQUENCE	3011 AA;	327142 MW; 772CBB29CCD94753 CRC64;

Query Match 89.5%; Score 137; DB 1; Length 3011;
 Best Local Similarity 89.3%; Pred. No. 2.8e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 PKPORKTKRAHRRPDQVKFPGGGGIVG 28
 ||||| :|||||
 Db 5 PKPORKTKRNTNRPRPDQVKFPGGGGIVG 32

RESULT 10
 POLG_HCVJ6
 ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.

P26660;
 01-AUG-1992 (Rel. 23, Created)
 01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 Hepatitis C virus (isolate HC-J6) (HCV).
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 NCBI_TaxID=11113;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=92044440; PubMed=1658196;
 Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.;
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 J. Gen. Virol. 72:2697-2704(1991).
 -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polypeptide, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 [RNA](N).
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@lsb-sib.ch).

 EMBL: D00944; BAA00792.1;
 PIR: J01303; J01303.
 HSSP: P27958; 1HEI.
 MEROPS: S29-001; -.
 MEROPS; U39-001; -.
 InterPro: IPR001410; DEAD.
 InterPro: IPR002522; HCV_capsid.
 InterPro: IPR002521; HCV_core.
 InterPro: IPR002519; HCV_env.
 InterPro: IPR002531; HCV_NS1.
 InterPro: IPR002518; HCV_NS2.
 InterPro: IPR004109; HCV_NS3.
 InterPro: IPR000745; HCV_NS4a.
 InterPro: IPR001490; HCV_NS4b.
 InterPro: IPR002868; HCV_NS5a.
 InterPro: IPR002166; HCV_RdRP.
 InterPro: IPR001650; Helicase_C.
 InterPro: IPR007095; RNA_pol_DS_PS.
 InterPro: IPR007094; RNA_pol_PSVir.
 Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.

DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Protease; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 89.5%; Score 137; DB 1; Length 3033;
 Best Local Similarity 89.3%; Pred. No. 2.9e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQRTKRAHRRPQDVKFGGQIVG 28
 Db 5 PKPQRTKRTNRPPQDVKFGGQIVG 32
 RESULT 11
 POLG_HCVJ8
 ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate HC-J6) (HCV).
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 DE Hepacivirus.
 NCBI_TaxID=11115;
 RN [1]

SEQUENCE FROM N.A.
 MEDLINE-92230232; PubMed-1314459;
 Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 Fukuda S., Tsuda F., Mishiro S.:
 "Full-length sequence of a hepatitis C virus genome having poor
 homology to reported isolates: comparative study of four distinct
 genotypes";
 Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10988; BAA01761.1;
 CC PIR: A40250; GNWJ08.
 CC HSP: P27958; 1HEV1.
 CC MEROPS: S29.001;
 CC MEROPS: U39.001;
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RDRP.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; Viral_RDRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDc; 1.
 CC PolyProtein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Nonstructural protein; Hydroxylase; Serine protease.
 CC INIT_MET 1
 CC 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 1010
 CC CHAIN 1011 1619
 CC CHAIN 1620 1866
 CC CHAIN 1867 2017
 CC CHAIN 2017

FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 239 239
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 FT SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381F01A CRC64;
 SQ
 Query Match 89.5%; Score 137; DB 1; Length 3033;
 Best Local Similarity 89.3%; Pred. No. 2.9e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKTKRKAKHRRPQDKVFPFGGQIVG 28
 DB 5 PKPQKTKRNTNRRPQDKVFPFGGQIVG 32
 RESULT 12
 POLG_HCVTW
 ID. POLG_HCVTW STANDARD; PRT: 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230206; PubMed-1314449;
 RA Chen P.-J., Lin M.H., Tai K.F., Liu P.-C., Lin C.-J., Chen D.-S.;
 RA "The Taiwanese hepatitis C virus genome: sequence determination and
 RA mapping the 5' termini of viral genome and antigenomic RNA";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M4754; ; NOT_ANNOTATED_CDS.

PIR; A40244; GNMVW.

PDB; 1N64; 25-FEB-93.

PDB; 1NS3; 08-APR-98.

MEROPS; S29.001; -.

InterPro; IPR001410; DEAD.

InterPro; IPR002522; HCV capsid.

InterPro; IPR002521; HCV core.

InterPro; IPR002519; HCV env.

InterPro; IPR002531; HCV NS1.

InterPro; IPR002518; HCV NS2.

InterPro; IPR004109; HCV NS3.

InterPro; IPR000745; HCV NS4a.

InterPro; IPR001490; HCV NS4b.

InterPro; IPR002868; HCV NS5a.

InterPro; IPR002166; HCV RdRP.

InterPro; IPR007095; RNA_pol_DS_PS.

InterPro; IPR007094; RNA_pol_psvir.

Pfam; PF01543; HCV capsid; 1.

Pfam; PF01542; HCV core; 1.

Pfam; PF01535; HCV env; 1.

Pfam; PF01560; HCV NS1; 1.

Pfam; PF01538; HCV NS2; 1.

Pfam; PF02907; HCV NS3; 1.

Pfam; PF01006; HCV NS4a; 1.

Pfam; PF01001; HCV NS4b; 1.

Pfam; PF01506; HCV NS5a; 1.

Pfam; PF00271; helicase_C; 1.

Pfam; PF00998; viral_RdRP; 1.

PRODOM; PD186062; HCV NS1; 1.

SMART; SM00487; DEXDc; 1.

Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

3D-structure.

INIT_MET 1 1

CHAIN 1 115

CHAIN 116 191

CHAIN 192 383

CHAIN 384 729

CHAIN 730 1006

CHAIN 1007 1615

CHAIN 1616 1862

CHAIN 1863 2013

CHAIN 2014 3010

CHAIN 3011 369

ACT_SITE 1083 1083

ACT_SITE 1107 1107

ACT_SITE 1165 1165

ACT_SITE 1230 1237

NP_BIND 1230 1319

SITE 1316 1319

CARBOHYD 196 196

CARBOHYD 209 209

CARBOHYD 233 233

CARBOHYD 234 234

CARBOHYD 235 235

CARBOHYD 305 305

CARBOHYD 417 417

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

CARBOHYD 423 423

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 85.0%; Score 130; DB 1; Length 3010;

Best Local Similarity 88.9%; Pred. No. 2.8e-10;

Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 KPQRKTRKAHRRPDVYKFGGGQIVG 28

Db 6 KPQRKTRKTRRRPDVYKFGGGQIVG 32

RESULT 13

POLG_HCV1

ID POLG_HCV1 STANDARD; PRT; 3011 AA.

AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

(GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)

(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]

OS Hepatitis C virus (isolate 1) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

Hepacivirus.

OC NCBI_TaxID=11104;

RP [1]

RN SEQUENCE FROM N.A.

RA MEDLINE-91172826; PubMed=1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,

Bradley D.W., Kuo G., Houghton M.

"Genetic organization and diversity of the hepatitis C virus.";

Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

-1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

precursor polyprotein, commonly with Asp or Glu in the p6

position, Cys or Thr in Pl and Ser or Ala in pl'.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

(RNA)(N).

-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

PROTEIN C AND MRNA.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M62321; AAA45676.1; -.

DR

DR PIR: A39166; GNWVC3.
 DR PDB: 1AIV; 16-FEB-99.
 DR PDB: 1HEI; 25-NOV-98.
 DR MEROPS: S29.001; -.
 DR MEROPS: U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01342; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 333
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1093
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789

SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 84.3%; Score 129; DB 1; Length 3011;
 Best Local Similarity 82.1%; Pred. No. 4e-10;
 Matches 23; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PKPQKTRKKAHRRPDVKFPGGGQIVG 28
 DB 5 PKPQKKNRNTNRRPDVKFPGGGQIVG 32
 RESULT 14
 IE63_HSV2H
 ID IE63_HSV2H STANDARD; PRT; 512 AA.
 AC P28276;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcriptional regulator IE63 (VMM63) (ICP27).
 GN UL54.
 OS Herpes simplex virus (type 2 / strain HG52).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxId=10315;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92113549; PubMed=1662697;
 RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
 RT "Comparative sequence analysis of the long repeat regions and
 RT adjoining parts of the long unique regions in the genomes of herpes
 RT simplex viruses types 1 and 2.";
 RL J. Gen. Virol. 72:3057-3075(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Dolan A.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBSJ databases.
 CC -!- FUNCTION: INVOLVED IN THE DOWN-REGULATION OF VIRAL IMMEDIATE-EARLY
 CC GENES AND ACTS IN COMBINATION WITH ICP0 AND ICP4 AS AN ACTIVATOR
 CC OF LATE GENES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
 CC HSV-2 UL54, EBV-1 5, VV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10471; BAA01269.1; -.
 DR EMBL; 286099; CAB06702.1; -.
 DR PIR: J01498; WMBEXA.
 KW Early protein; Transcription regulation; Activator; DNA-binding.
 SQ SEQUENCE 512 AA; 54958 MW; 459651470A503BA7 CRC64;
 Query Match 34.6%; Score 53; DB 1; Length 512;
 Best Local Similarity 37.5%; Pred. No. 4.6;
 Matches 12; Conservative 3; Mismatches 13; Indels 4; Gaps 1;
 QY 1 PKPQKTRKKAH---RRPDVKFPGGGQIVG 28
 DB 160 PKPRRVSNAHNOGRRHPASARTDGPATG 191
 RESULT 15
 FUR1_HUMAN
 ID FUR1_HUMAN STANDARD; PRT; 794 AA.
 AC P09958; Q14336;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Furin precursor (EC 3.4.21.75) (Paired basic amino acid residue

cleaving enzyme) (PACE) (Dibasic processing enzyme).

FURIN OR PACE OR FUR.

OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Blood;

RA MEDLINE=90175002; PubMed=2408021;

RX van den Ouweland A.M.W., van Duijnhoven H.L.P., Keizer G.D., Dorssers L.C.J., van de Ven W.J.M.;

RT "Structural homology between the human fur gene product and the subtilisin-like protease encoded by yeast KEX2.";

RL Nucleic Acids Res. 18:664-664(1990).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE=91321735; PubMed=1713771;

RX Barr P.J., Mason O.B., Landsberg K.E., Wong P.A., Kiefer M.C., Brake A.J.;

RA "cDNA and gene structure for a human subtilisin-like protease with cleavage specificity for paired basic amino acid residues.";

RL DNA Cell Biol. 10:319-328(1991).

RN [3]

RP SEQUENCE OF 296-794 FROM N.A.

RA MEDLINE=87053858; PubMed=3023061;

RX Roebroek A.J.M., Schaiken J.A., Leunissen J.A.M., Onnekink C., Bloemers H.P.J., van de Ven W.J.M.;

RA "Evolutionary conserved close linkage of the c-fes/fps proto-oncogene and genetic sequences encoding a receptor-like protein.";

RT EMBO J. 5:2197-2202(1986).

RL [4]

RP 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.

RA MEDLINE=94291619; PubMed=8020465;

RX Stezen R.J., Creemers J.W.M., van de Ven W.J.M.;

RA "Homology modelling of the catalytic domain of human furin. A model for the eukaryotic subtilisin-like proprotein convertases.";

RL Eur. J. Biochem. 222:255-266(1994).

RN [5]

RP PROCESSING.

RA MEDLINE=92332543; PubMed=1629222;

RX Leduc R., Molloy S.S., Thorne B.A., Thomas G.;

RA "Activation of human furin precursor processing endoprotease occurs by an intramolecular autoproteolytic cleavage.";

RL J. Biol. Chem. 267:14304-14308(1992).

CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTASE ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys. Releases albumin, complement component C3 and von Willebrand factor from their respective precursors.

CC -1- COFACTOR: CALCIUM-DEPENDENT.

CC -1- ENZYME REGULATION: COULD BE INHIBITED BY THE NOT SECONDLY CLEAVED PROPEPTIDE.

CC -1- SUBCELLULAR LOCATION: SEEMS TO BE LOCALIZED INTRACELLULARLY TO THE TRANS GOLGI NETWORK. PROPEPTIDE CLEAVAGE IS A PREREQUISITE FOR EXIT OF FURIN MOLECULES OUT OF THE ENDOPLASMIC RETICULUM (ER).

CC -1- SECOND CLEAVAGE IN THE PROPEPTIDE OCCUR IN THE TRANS GOLGI NETWORK (TGN), WHICH IS FOLLOWED BY THE RELEASE OF THE PROPEPTIDE BOUND TO FURIN AND THE ACTIVATION OF FURIN.

CC -1- TISSUE SPECIFICITY: SEEMS TO BE EXPRESSED UBIQUITOUSLY.

CC -1- DOMAIN: CONTAINS A CYTOPLASMIC DOMAIN RESPONSIBLE FOR ITS TGN LOCALIZATION AND RECYCLING FROM THE CELL SURFACE.

CC -1- PTM: THE PROPEPTIDE IS AUTOCATALYTICALLY REMOVED THROUGH AN INTRAMOLECULAR CLEAVAGE PROBABLY IN THE ENDOPLASMIC RETICULUM (ER). IN THE TGN THE SECOND CLEAVAGE IN THE PROPEPTIDE COULD LEAD TO THE ACTIVATION OF FURIN.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X17094; CAA34948.1; -

DR EMBL; X04329; CAA27860.1; -

DR EMBL; A06939; CAA00605.1; -

DR PIR; A39552; KXHUF.

DR HSSP; Q94405; LMPT.

DR MEROPS; S08.071; -

DR Genew; HGNC:8568; FURIN.

DR MIN; 136950; -

DR GO; GO:0005794; C:Golgi apparatus; TAS.

DR GO; GO:0004276; F:furin activity; TAS.

DR GO; GO:0007267; P:cell-cell signaling; TAS.

DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.

DR InterPro; IPR006212; Furin_repeat.

DR InterPro; IPR002884; P_domain.

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF01483; P_protein; PARTIAL.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P_domain; 1.

DR SMART; SM00261; FU; 2.

DR PROSITE; PS00136; SUBTILASE ASP; 1.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

DR Hydrolase; Serine protease; Transmembrane; Glycoprotein; Signal;

KW Zymogen; Calcium.

RN SIGNAL 1 24 POTENTIAL.

FT PROPEP 25 107

FT CHAIN 108 794 FURIN.

FT DOMAIN 556 705 CYS-RICH.

FT TRANSMEM 716 738 POTENTIAL.

FT ACT_SITE 153 153 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 194 194 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 211 360 POTENTIAL.

FT DISULFID 303 333 POTENTIAL.

FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SITE 70 75 CLEAVAGE (SECOND AUTO-).

FT SITE 104 107 CLEAVAGE (FIRST AUTO-).

FT SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 759 762 CELL SURFACE SIGNAL.

FT SITE 773 779 TRANS GOLGI NETWORK SIGNAL.

SQ SEQUENCE 794 AA; 86678 MW; 10C44DD5892EF85D CRC64;

Query Match 34.6%; Score 53; DB 1; Length 794;

Best Local Similarity 50.0%; Pred. No. 7.4;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 QKTKKRAHRRPQDVKEP 21
Db 102 KRTKRDVYQETDPKFP 119

Search completed: August 7, 2003, 11:20:06
Job time : 5.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-33

Perfect score: 153

Sequence: 1 PKPQRTKRAHRPQDVFPGGQIVG 28

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

tal number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvrius:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138	90.2	105	12 P90364	P90364 hepatitis c
2	138	90.2	106	12 Q81831	Q81831 hepatitis c
3	138	90.2	125	12 Q9PXN3	Q9PXN3 hepatitis c
4	138	90.2	191	12 Q68575	Q68575 hepatitis c
5	138	90.2	3010	12 Q91A00	Q91A00 hepatitis c
6	138	90.2	3010	12 Q81989	Q81989 hepatitis c
7	138	90.2	3010	12 Q81541	Q81541 hepatitis c
8	137	89.5	45	12 Q68311	Q68311 hepatitis c
9	137	89.5	45	12 Q68306	Q68306 hepatitis c
10	137	89.5	46	12 Q68309	Q68309 hepatitis c
11	137	89.5	60	12 Q8JYR9	Q8JYR9 hepatitis c
12	137	89.5	61	12 Q8JYR8	Q8JYR8 hepatitis c
13	137	89.5	61	12 Q8JYR5	Q8JYR5 hepatitis c
14	137	89.5	61	12 Q8JYR7	Q8JYR7 hepatitis c
15	137	89.5	62	12 Q8JYR6	Q8JYR6 hepatitis c
16	137	89.5	74	12 Q68712	Q68712 hepatitis c

17 137 89.5 74 12 Q68682 hepatitis c
18 137 89.5 74 12 Q68708 hepatitis c
19 137 89.5 74 12 Q68679 hepatitis c
20 137 89.5 74 12 Q68683 hepatitis c
21 137 89.5 74 12 Q68706 hepatitis c
22 137 89.5 74 12 Q68684 hepatitis c
23 137 89.5 74 12 Q68707 hepatitis c
24 137 89.5 74 12 Q68685 hepatitis c
25 137 89.5 76 12 Q8JYS0 hepatitis c
26 137 89.5 78 12 Q8JYS1 hepatitis c
27 137 89.5 83 12 Q81264 hepatitis c
28 137 89.5 100 12 Q9QT61 hepatitis c
29 137 89.5 100 12 Q8QP70 hepatitis c
30 137 89.5 100 12 Q8QP54 hepatitis c
31 137 89.5 100 12 Q9QT57 hepatitis c
32 137 89.5 100 12 Q8QP52 hepatitis c
33 137 89.5 100 12 Q8QP88 hepatitis c
34 137 89.5 100 12 Q9QT56 hepatitis c
35 137 89.5 100 12 Q8QP90 hepatitis c
36 137 89.5 100 12 Q8QP64 hepatitis c
37 137 89.5 100 12 Q8QP55 hepatitis c
38 137 89.5 100 12 Q8QP60 hepatitis c
39 137 89.5 100 12 Q8QP57 hepatitis c
40 137 89.5 100 12 Q9QT58 hepatitis c
41 137 89.5 100 12 Q8QP63 hepatitis c
42 137 89.5 100 12 Q8QP61 hepatitis c
43 137 89.5 100 12 Q9QT52 hepatitis c
44 137 89.5 100 12 Q9QT54 hepatitis c
45 137 89.5 100 12 Q8QP59 hepatitis c

ALIGNMENTS

RESULT 1

P90364 ID P90364 PRELIMINARY; PRT; 105 AA.
AC P90364;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polypeptide) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohno T., Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR method."
RL Thesis (1995), Nagoya City University Medical School.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305314; PubMed=8712927;
RA Ohno T., Mizokami M., Saleh M.G., Orito E., Ohba K.-I., Wu R.-R., Koide T., Tibbs C.J., Nouri-Aria K.T., Tokudome S., Williams R.;
RT "Usefulness and limitation of phylogenetic analysis for hepatitis C virus core region: application to isolates from Egyptian and Yemeni patients."
RL Arch. Virol. 141:1101-1113(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; D82034; BAA11519.1; -;
DR InterPro; IPR002522; HCV_capsid.
KW Pfam; PF01543; HCV_capsid; 1.
FT polyprotein.
FT NON_TER 1 105
FT TER 105
SQ SEQUENCE 105 AA; 11870 MW; C7BA40B284025A49 CRC64;


```
Query Match          90.2%; Score 138; DB 12; Length 105;
Best Local Similarity 89.3%; Pred. No. 8.7e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
DB 4 PAPQKTKRNTYRRPDQVKFPGGGQIVG 31

RESULT 2
Q81831 PRELIMINARY; PRT; 106 AA.
AC Q81831;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE (isolate USA8) genomic RNA (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN NCBI_TaxID=111103;
[1]
RC SEQUENCE FROM N.A.
RC STRAIN-USAB;
RA Ohno T., Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR
RT method.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D49465; BAA08439.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 106 AA; 12001 MW; 25D0D5414B3E9A9DC CRC64;

Query Match          90.2%; Score 138; DB 12; Length 106;
Best Local Similarity 89.3%; Pred. No. 8.8e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 3
XN3
Q9PXN3 PRELIMINARY; PRT; 125 AA.
AC Q9PXN3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN NCBI_TaxID=40271;
[1]
RC SEQUENCE FROM N.A.
RC MEDLINE=94351179; PubMed=7520922;
RA Nakazawa T., Kato N., Ohkoshi S., Shibuya A., Shimotohno K.;
RL J. Hepatol. 20:623-629(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
SQ SEQUENCE 125 AA; 14220 MW; D945CB60661797E3 CRC64;
```

```
Query Match          90.2%; Score 138; DB 12; Length 125;
Best Local Similarity 89.3%; Pred. No. 1e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 4
Q88575 PRELIMINARY; PRT; 191 AA.
AC Q88575;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN NCBI_TaxID=111103;
[1]
RC SEQUENCE FROM N.A.
RC STRAIN-BB51;
RA Songvilalai S., Dharakul T., Kunkitti R., Thepthai C.;
RT "Molecular cloning and expression of hepatitis C virus core protein
RT and production of monoclonal antibodies to the recombinant protein.";
RL Asian Pac. J. Allergy Immunol. 14:0-0(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; U55284; AAB00215.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 191 AA; 20840 MW; 4AAE63444D8329E2 CRC64;

Query Match          90.2%; Score 138; DB 12; Length 191;
Best Local Similarity 89.3%; Pred. No. 1.6e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 5
Q91AU0 PRELIMINARY; PRT; 3010 AA.
AC Q91AU0;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN NCBI_TaxID=111103;
[1]
RC SEQUENCE FROM N.A.
RC STRAIN-HCV-S1;
RX MEDLINE=21440119; PubMed=11556407;
RA Lim S.P., Khu Y.L., Hong W.J., Tay A., Ting A.E., Lim S.G., Tan Y.H.;
RT "Identification and molecular characterization of the complete genome
RT of a Singapore isolate of hepatitis C virus: sequence comparison with
RT other strains and phylogenetic analysis.";
RL Virus Genes 23:89-95(2001).
RN [2]
```

RP SEQUENCE FROM N.A.
RC Lim S.P.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF356827; AAL00900.1; -
DR InterPro; IPR000345; CytC_hemo_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; transferase; transmembrane.
SQ SEQUENCE 3010 AA; 326793 MW; 3D9304314F9F795 CRC64;
Query Match 90.2%; Score 138; DB 12; Length 3010;
Best Local Similarity 89.3%; Pred. No. 3e-11;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Oy 1 PKPQKTKRKAHRRPQDVKPGGGQIVG 28
||||||| :|||||||
5 PKPQKTKRNTYRRPQDVKPGGGQIVG 32
RESULT 6
Q81989 PRELIMINARY; PRT; 3010 AA.
AC Q81989;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E1 and E2/NS1 envelope glycoprotein (Genome polyprotein).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-N;
RA Zheng W.-Y.Z.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-N;
RA Zheng W.Z.;
RT "Genotype identification of hepatitis c virus (HCV) isolated from a

RT single Japanese carrier in Nagasaki prefecture and genome analysis of
RT E1 and E2/NS1 envelope glycoprotein regions.*;
RL Jpn. J. Trop. Med. Hyg. 22:169-177(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D63857; BAA09919.1; -
DR HSSP; P26663; LUXP.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327503 MW; C7BDB38169D6E3CF CRC64;
Query Match 90.2%; Score 138; DB 12; Length 3010;
Best Local Similarity 89.3%; Pred. No. 3e-11;
Matches 25; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Oy 1 PKPQKTKRKAHRRPQDVKPGGGQIVG 28
||||||| :|||||||
5 PKPQKTKRNTYRRPQDVKPGGGQIVG 32
Db
RESULT 7
Q81541 PRELIMINARY; PRT; 3010 AA.
AC Q81541;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JT';
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T.,
RA Hijikata M., Ishimura Y., Shimotohno K.;
RA "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.*;

Virus Res. 23:39-53(1992).

-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

EMBL: D11355; BAA18894.1; -

HSPF; P26663; LXP.

InterPro: IPR001410; DEAD.

InterPro: IPR002522; HCV_capsid.

InterPro: IPR002521; HCV_core.

InterPro: IPR002519; HCV_env.

InterPro: IPR002531; HCV_NS1.

InterPro: IPR002518; HCV_NS2.

InterPro: IPR004109; HCV_NS3.

InterPro: IPR000745; HCV_NS4a.

InterPro: IPR001490; HCV_NS4b.

InterPro: IPR002868; HCV_NS5a.

InterPro: IPR002166; HCV_RdRP.

InterPro: IPR007095; RNA_pol_DS_PS.

InterPro: IPR007094; RNA_pol_PSVir.

Pfam: PF01543; HCV_capsid; 1.

Pfam: PF01542; HCV_core; 1.

Pfam: PF01539; HCV_env; 1.

Pfam: PF01560; HCV_NS1; 1.

Pfam: PF01538; HCV_NS2; 1.

Pfam: PF02907; HCV_NS3; 1.

Pfam: PF01006; HCV_NS4a; 1.

Pfam: PF01001; HCV_NS4b; 1.

Pfam: PF01506; HCV_NS5a; 1.

Pfam: PF00998; Viral_RdRP; 1.

ProDom: PD186062; HCV_NS1; 1.

SMART: SM00487; DEXDC; 1.

PROSITE: PS05057; RDRP_POSITIVE; 1.

PROSITE: PS05051; RDRP_VIRAL; 1.

Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.

FT CHAIN 1 191

FT CHAIN 192 383

FT CHAIN 384 729

FT CHAIN 730 1006

FT CHAIN 1007 1615

FT CHAIN 1616 1862

FT CHAIN 1863 2013

FT CHAIN 2014 3010

FT CHAIN 3010 3010

SEQUENCE 3010 AA; 326564 MW; 05F0B2102CF9DD9D CRC64;

Query Match 90.2%; Score 138; DB 12; Length 3010;

Best Local Similarity 89.3%; Pred. No. 3e-11; Mismatches 1; Indels 0; Gaps 0;

atches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 PKPQRTKTKRAHRRPDVKKFPGGGQIVG 28

||||||| :|||||||

5 PKPQRTKTKRNTNRRPDVKKFPGGGQIVG 32

RESULT 8

Q68311

ID Q68311 PRELIMINARY; PRT; 45 AA.

AC Q68311;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RC STRAIN-HCV-BB37;

RA Songsivilai S., Kanistanon D., Kunkitti R.;

RT "Identification and characterisation of Thai isolates of hepatitis C virus."

Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

EMBL: U23749; AAA65056.1; -

InterPro: IPR002522; HCV_capsid.

Pfam: PF01543; HCV_capsid; 1.

Polyprotein. 45 45

NON_TER

SEQUENCE 45 AA; 5015 MW; CC527167096AAA81 CRC64;

Query Match 89.5%; Score 137; DB 12; Length 45;

Best Local Similarity 89.3%; Pred. No. 5e-13; Mismatches 1; Indels 0; Gaps 0;

atches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 PKPQRTKTKRAHRRPDVKKFPGGGQIVG 28

||||||| :|||||||

5 PKPQRTKTKRNTNRRPDVKKFPGGGQIVG 32

RESULT 9

Q68306

ID Q68306 PRELIMINARY; PRT; 45 AA.

AC Q68306;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RC STRAIN-HCV-BB5;

RA Songsivilai S., Kanistanon D., Kunkitti R.;

RT "Identification and characterisation of Thai isolates of hepatitis C virus."

Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

EMBL: U23744; AAA65051.1; -

InterPro: IPR002522; HCV_capsid.

Pfam: PF01543; HCV_capsid; 1.

Polyprotein. 45 45

NON_TER

SEQUENCE 45 AA; 5030 MW; CC4C21ED236AAA81 CRC64;

Query Match 89.5%; Score 137; DB 12; Length 45;

Best Local Similarity 89.3%; Pred. No. 5e-13; Mismatches 1; Indels 0; Gaps 0;

atches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 PKPQRTKTKRAHRRPDVKKFPGGGQIVG 28

||||||| :|||||||

5 PKPQRTKTKRNTNRRPDVKKFPGGGQIVG 32

RESULT 10

Q68309

ID Q68309 PRELIMINARY; PRT; 46 AA.

AC Q68309;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RC STRAIN-HCV-BB37;

RA Songsivilai S., Kanistanon D., Kunkitti R.;

RT "Identification and characterisation of Thai isolates of hepatitis C virus."

```

RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB10;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT *Identification and characterisation of Thai isolates of hepatitis C
RT virus.*;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; U23747; AA65054.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 46 AA; 5129 MW; ABD4C421ED36AAA CRC64;

Query Match
Best Local Similarity 89.5%; Score 137; DB 12; Length 46;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 11
QBRYR9 PRELIMINARY; PRT; 60 AA.
AC QBRYR9;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; AF506616; AA633392.1; -.
InterPro; IPR002522; HCV_capsid.
Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 60 AA; 6686 MW; B5691CBE8F26F2F1 CRC64;

Query Match
Best Local Similarity 89.5%; Score 137; DB 12; Length 60;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 12
QBRYR8 PRELIMINARY; PRT; 61 AA.
AC QBRYR8;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.

```

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG286;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.*;
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF506617; AA633393.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match
Best Local Similarity 89.5%; Score 137; DB 12; Length 61;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

Db 5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 13
QBRYR5 PRELIMINARY; PRT; 61 AA.
AC QBRYR5;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF506620; AA633396.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match
Best Local Similarity 89.5%; Score 137; DB 12; Length 61;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

Db 5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 14
QBRYR7 PRELIMINARY; PRT; 61 AA.
AC QBRYR7;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

```

01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
Genome polyprotein (Fragment).
Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
NCBI_TaxID=11103;
[1]
SEQUENCE FROM N.A.
STRAIN=RIG289;
Shustov A.V., Gavrilova I.V., Netesov S.V.;
"Genetic variability of hepatitis C virus in Western Siberia.";
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; AF506618; AAM33394.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
DR Polyprotein.
NON_TER 61 61
SEQUENCE 61 AA: 6869 MW: 66577068E8F26F2 CRC64;
Query Match 89.5%; Score 137; DB 12; Length 61;
Best Local Similarity 89.3%; Pred. No. 6.9e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 PKPQRTKTKAHRRPDQVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKTKAHRRPDQVKFPGGGQIVG 32
RESULT 15
Q8JYR6 PRELIMINARY; PRT; 62 AA.
AC Q8JYR6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIG289;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
"Genetic variability of hepatitis C virus in Western Siberia.";
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; AF506619; AAM33395.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
DR Polyprotein.
FT NON_TER 62 62
SQ SEQUENCE 62 AA: 7053 MW: EEE656DC79E8F26 CRC64;
Query Match 89.5%; Score 137; DB 12; Length 62;
Best Local Similarity 89.3%; Pred. No. 7e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 PKPQRTKTKAHRRPDQVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PAPQRTKTKAHRRPDQVKFPGGGQIVG 32
Search completed: August 7, 2003, 11:19:03
Job time : 25.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-33

Perfect score: 153
Sequence: 1 PXPQKTKRAHRRPDVRFPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

tal number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/6CTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	153	100.0	28	3	US-08-921-887-33
2	148	96.7	28	3	US-08-921-887-28
3	138	90.2	28	3	US-08-921-887-25
4	138	90.2	450	4	US-08-635-886C-191
5	138	90.2	450	4	US-08-635-886C-192
6	137	89.5	34	3	US-08-380-160-6
7	137	89.5	43	4	US-09-020-846-36
8	137	89.5	44	3	US-08-380-160-2
9	137	89.5	44	4	US-09-389-756-1
10	137	89.5	45	3	US-08-380-160-1
11	137	89.5	61	1	US-07-946-054-9
12	137	89.5	61	1	US-08-947-23
13	137	89.5	61	1	US-08-530-550-3
14	137	89.5	61	1	US-08-262-037-26
15	137	89.5	61	5	PCT-US93-08638-9
16	137	89.5	61	5	PCT-US94-07088-23
17	137	89.5	61	5	PCT-US95-13660-3
18	137	89.5	74	3	US-08-836-075A-10
19	137	89.5	74	4	US-08-635-886C-198
20	137	89.5	100	4	US-08-635-886C-232
21	137	89.5	108	3	US-08-836-075A-14
22	137	89.5	115	1	US-08-324-977-8
23	137	89.5	115	2	US-08-384-616-8
24	137	89.5	115	2	US-08-904-686A-8
25	137	89.5	115	3	US-09-315-850-8
26	137	89.5	123	2	US-08-501-195-2
27	137	89.5	137	3	US-08-836-075A-46

28	137	89.5	138	3	US-08-836-075A-60
29	137	89.5	140	2	US-08-501-195-1
30	137	89.5	154	3	US-08-854-531-2
31	137	89.5	154	5	PCT-US95-13552-2
32	137	89.5	190	1	US-07-681-701-16
33	137	89.5	190	3	US-08-078-271B-1
34	137	89.5	191	2	US-08-290-665A-155
35	137	89.5	191	2	US-08-290-665A-156
36	137	89.5	191	2	US-08-290-665A-157
37	137	89.5	191	2	US-08-290-665A-158
38	137	89.5	191	2	US-08-290-665A-159
39	137	89.5	191	2	US-08-290-665A-160
40	137	89.5	191	2	US-08-290-665A-161
41	137	89.5	191	2	US-08-290-665A-163
42	137	89.5	191	2	US-08-290-665A-164
43	137	89.5	191	2	US-08-290-665A-165
44	137	89.5	191	2	US-08-290-665A-166
45	137	89.5	191	2	US-08-290-665A-167

ALIGNMENTS

RESULT 1
US-08-921-887-33
: Sequence 33, Application US/08921887
: Patent No. 6030771
: GENERAL INFORMATION:
: APPLICANT: KHUDYAKOV, YURI E.
: TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAM
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: JONES & ASKEW, LLP
: STREET: 191 Peachtree Street, N.W., 37th Floor
: CITY: Atlanta
: STATE: GA
: COUNTRY: USA
: ZIP: 30303-1769
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/921,887
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: WARREN, WILLIAM L.
: REGISTRATION NUMBER: 36,714
: REFERENCE/DOCKET NUMBER: 03063-0380
: TELEPHONE: 404-818-3700
: TELEFAX: 404-818-3799
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: ORGANISM: Hepatitis virus
US-08-921-887-33

Query Match 100.0%; Score 153; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 5.8e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGQIVG 28
|||||
DB 1 PKPQKTKRKAHRRPDQVKFPGGQIVG 28

RESULT 2

US-08-921-887-28
; Sequence 28, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION.
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus

US-08-921-887-28
; Sequence 28, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:

RESULT 3

US-08-921-887-25
; Sequence 25, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:

RESULT 4

US-08-921-887-25
; Sequence 25, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:

ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus

US-08-921-887-25

US-08-921-887-25
; Sequence 25, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-921-887-25

US-08-921-887-25
; Sequence 25, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-921-887-25

US-08-921-887-25
; Sequence 25, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-921-887-25

US-08-921-887-25
; Sequence 25, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-921-887-25

US-08-921-887-25
; Sequence 25, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-921-887-25

US-08-921-887-25
; Sequence 25, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus

US-08-921-887-25

QY 1 PKPQKTRKKAHRRPQDVKFGGGOIVG 28
||||| :|||||
Db 5 PKPQKTRKTRTYRRPQDVKFGGGOIVG 32

RESULT 5

US-08-635-886C-192
; Sequence 192, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 192
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-192

Query Match 90.2%; Score 138; DB 4; Length 450;

Best Local Similarity 89.3%; Pred. No. 1.2e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTRKKAHRRPQDVKFGGGOIVG 28
||||| :|||||
Db 5 PKPQKTRKTRTYRRPQDVKFGGGOIVG 32

RESULT 6

US-08-380-160-6
; Sequence 6, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380.160
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/057,471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 28682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-8400
; TELEFAX: (703)836-2787
; TELEX:

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
US-08-380-160-6

Query Match 89.5%; Score 137; DB 3; Length 34;

Best Local Similarity 89.3%; Pred. No. 1.2e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTRKKAHRRPQDVKFGGGOIVG 28
||||| :|||||
Db 4 PKPQKTRKTRTYRRPQDVKFGGGOIVG 31

RESULT 7

US-09-020-846-36
; Sequence 36, Application US/09020846
; Patent No. 6322965
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, Kenjiro
; APPLICANT: KASHIWAKUMA, Tomiko
; APPLICANT: CHIBA, Yukie
; APPLICANT: YAGI, Shintaro
; APPLICANT: HASEGAWA, Akira
; TITLE OF INVENTION: CHIMERA ANTIGEN PEPTIDE
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,846
; FILING DATE: 09-FEB-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-027015
; FILING DATE: 10-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-024045
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 053466/0225
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-020-846-36

Query Match 89.5%; Score 137; DB 4; Length 43;
Best Local Similarity 89.3%; Pred. No. 1.5e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 8
US-08-380-160-2
; Sequence 2, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/380,160
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,471
FILING DATE: 06-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 28682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-6400
TELEFAX: (703)836-2787
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Human Hepatitis C Virus
STRAIN: H77
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..44
OTHER INFORMATION: /note= "N-terminal sequence of the
OTHER INFORMATION: protein of the nucleocapside or CORE protein of
OTHER INFORMATION: the human hepatitis C virus"
US-08-380-160-2

Query Match 89.5%; Score 137; DB 3; Length 44;
Best Local Similarity 89.3%; Pred. No. 1.6e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQKTKRNTNRRPDQVKFPGGGQIVG 31

RESULT 9
US-09-389-756-1
; Sequence 1, Application US/09389756
; Patent No. 6576240
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/09/389,756
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: PCT/FR98/00442
; EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44

TYPE: PRT
ORGANISM: Hepatitis C virus
PUBLICATION INFORMATION:
AUTHORS: Ogata, N. et al.
TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
Patent No. 6576240
TITLE: of Hepatitis Virus
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 3392-3396
DATE: 1991
RELEVANT RESIDUES: 2 TO 45
US-09-389-756-1

Query Match 89.5%; Score 137; DB 4; Length 44;
Best Local Similarity 89.3%; Pred. No. 1.6e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDQVKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQKTKRNTNRRPDQVKFPGGGQIVG 31

RESULT 10
US-08-380-160-1
; Sequence 1, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320

COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/380,160
;; FILING DATE:
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/057,471
;; FILING DATE: 06-MAY-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berridge, William P.
;; REGISTRATION NUMBER: 30,024
;; REFERENCE/DOCKET NUMBER: WPB 28682
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703)836-6400
;; TELEFAX: (703)836-2787
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 45 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;; ORIGINAL SOURCE:
;; ORGANISM: Human Hepatitis C Virus
;; STRAIN: H77
;; FEATURE:
;; NAME/KEY: Peptide
;; LOCATION: 1..45
;; OTHER INFORMATION: /note= "N-terminal sequence of the
;; OTHER INFORMATION: protein of the nucleocapsid or CORE protein of
;; OTHER INFORMATION: the human hepatitis C virus"
US-08-380-160-1

Query Match 89.5%; Score 137; DB 3; Length 45;
Best Local Similarity 89.3%; Pred. No. 1.6e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTRKRAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQKTRKNTNRRPDVKFPGGGQIVG 32

ULT 11
US-07-946-054-9
;; Sequence 9, Application US/07946054
;; Patent No. 5582968
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Chang Y1
;; APPLICANT: Hosein, Barbara H
;; TITLE OF INVENTION: No. 5582968el Branched Hybrid and Cluster
;; TITLE OF INVENTION: Peptides Effective in Diagnosing and Detecting No. 5582968-A,
;; TITLE OF INVENTION: No. 5582968-B Hepatitis
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: United Biomedical Inc.
;; STREET: 25 Davids Dr.
;; CITY: Hauppauge
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 11788
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/07/946,054
;; FILING DATE: 15-SEP-1992
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wilson, M. Lisa
;; REGISTRATION NUMBER: 34,045
;; REFERENCE/DOCKET NUMBER: 2000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 516-273-1717
;; TELEFAX: 516-273-2828
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 61 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-07-946-054-9

Query Match 89.5%; Score 137; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTRKRAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQKTRKNTNRRPDVKFPGGGQIVG 31

RESULT 12
US-08-083-947-23
;; Sequence 23, Application US/08083947
;; Patent No. 5639594
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Chang Y1
;; APPLICANT: Hosein, Barbara
;; TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
;; TITLE OF INVENTION: in Diagnosing and Detecting No. 5639594-A, No. 5639594-1
;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: M. Lisa Wilson
;; STREET: 25 Davids Drive
;; CITY: Hauppauge
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 11788
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/083,947
;; FILING DATE: 19930628
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 946,054
;; FILING DATE: 15-SEP-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wilson, M. Lisa
;; REGISTRATION NUMBER: 34045
;; REFERENCE/DOCKET NUMBER: 2000Z
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (516)273-2828
;; TELEFAX: (516)273-1717
;; INFORMATION FOR SEQ ID NO: 23:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 61 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-083-947-23

Query Match 89.5%; Score 137; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRKAHRRPDQVKKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQRTKTRNTNRRPDQVKKFPGGGQIVG 31

RESULT 13
US-08-530-550-3
; Sequence 3, Application US/08530550
; Patent No. 5736321
; GENERAL INFORMATION:
; APPLICANT: Hosein, Barbara
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Peptides Effective for Diagnosis and
; TITLE OF INVENTION: Detection of Hepatitis c Infection
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,550
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 2000Z
; TELEPHONE: (516)273-2828
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-530-550-3

Query Match 89.5%; Score 137; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRKAHRRPDQVKKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQRTKTRNTNRRPDQVKKFPGGGQIVG 31

RESULT 14
US-08-262-037-26
; Sequence 26, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA

ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
US-08-262-037-26

Query Match 89.5%; Score 137; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKTKRKAHRRPDQVKKFPGGGQIVG 28
||||||| :|||||||
Db 4 PKPQRTKTRNTNRRPDQVKKFPGGGQIVG 31

RESULT 15
PCT-US93-08638-9
; Sequence 9, Application PC/TUS9308638
; GENERAL INFORMATION:
; APPLICANT: United Biomedical Inc.
; TITLE OF INVENTION: Novel Branched Hybrid and Cluster Peptides
; TITLE OF INVENTION: Effective in Diagnosing and Detecting Non-A,
; TITLE OF INVENTION: Non-B Hepatitis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: UNITED BIOMEDICAL INC.
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: New York
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08638
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Lisa Wilson

```

; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 9055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-2828
; TELEFAX: 516-273-1717
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-08638-9

```

```

Query Match      89.5%; Score 137; DB 5; Length 61;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY      1 PKPQKTKRKARRPDVKFPGGGQIVG 28
          ||||| :|||||
          4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31

```

Search completed: August 7, 2003, 11:23:53
Job time : 10.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146A-33

Perfect score: 153

Sequence: 1 PKQPKYKRAHRRPQDKFPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

al number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PublishedApplications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137	89.5	44	15	US-10-367-677-1
2	137	89.5	74	10	Sequence 1, Appli
3	137	89.5	91	9	Sequence 10, Appl
4	137	89.5	97	10	Sequence 1, Appli
5	137	89.5	103	10	Sequence 77, Appl
6	137	89.5	108	10	Sequence 14, Appl
7	137	89.5	113	10	Sequence 78, Appl
8	137	89.5	137	10	Sequence 46, Appl
9	137	89.5	138	10	Sequence 60, Appl
10	137	89.5	166	11	Sequence 152, App
11	137	89.5	166	11	Sequence 152, App
12	137	89.5	169	11	Sequence 42, Appl
13	137	89.5	169	11	Sequence 42, Appl
14	137	89.5	169	11	Sequence 42, Appl
15	137	89.5	169	11	Sequence 44, Appl

Sequence 2, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 76, Appli
Sequence 12, Appli
Sequence 18, Appli
Sequence 50, Appli
Sequence 52, Appli
Sequence 54, Appli
Sequence 144, App
Sequence 144, App
Sequence 50, Appli
Sequence 52, Appli
Sequence 54, Appli
Sequence 144, App
Sequence 50, Appli
Sequence 50, Appli
Sequence 50, Appli
Sequence 23, Appli
Sequence 23, Appli
Sequence 40, Appli
Sequence 4, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 2, Appli
Sequence 2, Appli

16 137 89.5 182 10 US-09-929-955-2
17 137 89.5 182 14 US-10-104-966-2
18 137 89.5 191 11 US-09-194-949-3
19 137 89.5 318 10 US-09-851-138-76
20 137 89.5 319 10 US-09-851-138-12
21 137 89.5 319 10 US-09-851-138-18
22 137 89.5 319 11 US-09-899-046-50
23 137 89.5 319 11 US-09-899-046-52
24 137 89.5 319 11 US-09-899-046-54
25 137 89.5 319 11 US-09-899-046-144
26 137 89.5 319 11 US-09-878-281-50
27 137 89.5 319 11 US-09-878-281-52
28 137 89.5 319 11 US-09-878-281-54
29 137 89.5 319 11 US-09-878-281-144
30 137 89.5 809 10 US-09-973-025-50
31 137 89.5 809 11 US-09-899-303-50
32 137 89.5 809 11 US-09-995-808-50
33 137 89.5 809 11 US-09-995-860-50
34 137 89.5 2894 10 US-09-941-611-23
35 137 89.5 2894 15 US-10-044-995-23
36 137 89.5 2985 15 US-10-259-275-40
37 137 89.5 3011 9 US-09-742-659-4
38 137 89.5 3011 10 US-09-952-572-9
39 137 89.5 3011 10 US-09-929-955-1
40 137 89.5 3011 10 US-09-747-419-20
41 137 89.5 3011 11 US-09-891-894-3
42 137 89.5 3011 14 US-10-104-966-1
43 137 89.5 3011 15 US-10-259-275-20
44 137 89.5 3012 10 US-09-238-076-2
45 137 89.5 3012 11 US-09-995-937-2

ALIGNMENTS

RESULT 1
US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND TREATING AN HCV INFECTION
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10367.677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLIC INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; TITLE: Of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match 89.5% Score 137; DB 15; Length 44;

Best Local Similarity 89.3%; Pred. No. 1.2e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQKTKRKAHRRPDVKFPGGGQIVG 28
Db 4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31

RESULT 2

US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Query Match 89.5%; Score 137; DB 10; Length 74;
Best Local Similarity 89.3%; Pred. No. 2.1e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDVKFPGGGQIVG 28
Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 3

US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US20020090607A1
; GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS
; FILE REFERENCE: 14114.034902
; CURRENT APPLICATION NUMBER: US/09/758,308

; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 89.5%; Score 137; DB 9; Length 91;
Best Local Similarity 89.3%; Pred. No. 2.6e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDVKFPGGGQIVG 28
Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 4

US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 89.5%; Score 137; DB 10; Length 97;
Best Local Similarity 89.3%; Pred. No. 2.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRKAHRRPDVKFPGGGQIVG 28
Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 5

US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Hepatitis C virus
--09-921-397-77
Query Match 89.5%; Score 137; DB 10; Length 103;
Best Local Similarity 89.3%; Pred. No. 2.9e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQKTKRKAHRRPQDVKPPGGQIVG 28
Db 18 PKPQKTKRNTNRRPQDVKPPGGQIVG 45
RESULT 6
US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851.138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-09-851-138-14
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
Query Match 89.5%; Score 137; DB 10; Length 108;
Best Local Similarity 89.3%; Pred. No. 3.1e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQKTKRKAHRRPQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32
RESULT 7
US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-78
Query Match 89.5%; Score 137; DB 10; Length 113;
Best Local Similarity 89.3%; Pred. No. 3.2e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQKTKRKAHRRPQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32
RESULT 8
US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851.138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```

; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46
Query Match      89.5%; Score 137; DB 10; Length 137;
Best Local Similarity 89.3%; Pred. No. 3.9e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32
RESULT 9
US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-851-138-60
Query Match      89.5%; Score 137; DB 10; Length 138;
Best Local Similarity 89.3%; Pred. No. 3.9e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32
US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-152
Query Match      89.5%; Score 137; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 4.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32
US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-152
Query Match      89.5%; Score 137; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 4.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32

```

```

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32
RESULT 10
US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-152
Query Match      89.5%; Score 137; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 4.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32
RESULT 11
US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-152
Query Match      89.5%; Score 137; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 4.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKTKRAHRRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRTNRRPDQVKFPGGGQIVG 32

```


Best Local Similarity 89.3%; Pred. No. 4.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKRKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKRNTNRRPDVKFPGGGQIVG 32

RESULT 12

US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-42

Query Match 89.5%; Score 137; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 4.8e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKRNTNRRPDVKFPGGGQIVG 32

RESULT 13

US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-44

Query Match 89.5%; Score 137; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 4.8e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKRKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKRNTNRRPDVKFPGGGQIVG 32

RESULT 14

US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-42

Query Match 89.5%; Score 137; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 4.8e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKRNTNRRPDVKFPGGGQIVG 32

RESULT 15

US-09-878-281-44
; Sequence 44, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-44

Best Local Similarity 89.3%; Pred. No. 4.7e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKRKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKRNTNRRPDVKFPGGGQIVG 32

RESULT 12

US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-42

Query Match 89.5%; Score 137; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 4.8e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRKAHRRPDVKFPGGGQIVG 28
||||||| :|||||||
Db 5 PKPQRTKRNTNRRPDVKFPGGGQIVG 32

RESULT 13

US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-44

Query Match 89.5%; Score 137; DB 11; Length 169;
 Best Local Similarity 89.3%; Pred. No. 4.8e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 PKPQKTKRKAHRRPQDVKPPGGQIVG 28
 Db 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32

Search completed: August 7, 2003, 12:01:15
 Job time : 15.3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds
(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146A-32

Perfect score: 149

Sequence: 1 PKPKRQTKRNTLRPNKVKFPAGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues 1107863
al number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	149	100.0	28	20	AA1980682	NC mosaic protein
2	149	100.0	3023	17	AA1984462	Hepatitis C virus
3	130	87.2	470	14	AA1934473	Encoded by Hepati
4	130	87.2	470	14	AA1934474	Encoded by Hepati
5	130	87.2	470	14	AA1934475	Encoded by Hepati
6	126	84.6	28	20	AA1986673	NC mosaic protein
7	126	84.6	189	23	AA191258	HCV type 3 capsid
8	126	84.6	191	17	AA192968	Hepatitis C virus
9	126	84.6	191	17	AA192969	Hepatitis C virus

10	126	84.6	191	17	AA1989	Hepatitis C virus
11	126	84.6	319	17	AA1990	Hepatitis C virus
12	125	83.9	28	20	AA1991	NC mosaic protein
13	125	83.9	74	17	AA1992	Hepatitis C virus
14	125	83.9	191	17	AA1993	Hepatitis C virus
15	125	83.9	191	17	AA1994	Hepatitis C virus
16	125	83.9	470	14	AA1995	Encoded by Hepatit
17	123	82.6	3010	15	AA1996	NC mosaic protein
18	123	82.6	3010	23	AA1997	Blood transmissibl
19	123	82.6	3010	23	AA1998	HCV-S1 full-length
20	122	81.9	40	16	AA1999	Hepatitis C virus
21	122	81.9	40	16	AA2000	Hepatitis C virus
22	122	81.9	50	16	AA2001	Hepatitis C virus
23	122	81.9	82	13	AA2002	Hepatitis C virus
24	122	81.9	82	13	AA2003	Non-A, Non-B Hepat
25	122	81.9	120	18	AA2004	Hepatitis C virus
26	122	81.9	150	18	AA2005	Hepatitis C virus
27	122	81.9	176	18	AA2006	Mouse protamine 1
28	122	81.9	191	17	AA2007	Hepatitis C virus
29	122	81.9	191	17	AA2008	Recombinant protei
30	121	81.2	36	16	AA2009	CN14 fragment of H
31	121	81.2	38	14	AA2010	HCV capsid peptide
32	121	81.2	38	14	AA2011	HCV capsid peptide
33	121	81.2	38	14	AA2012	HCV capsid peptide
34	121	81.2	38	15	AA2013	Non-A, non-B hepat
35	121	81.2	43	19	AA2014	Hepatitis C virus
36	121	81.2	44	19	AA2015	Hepatitis C virus
37	121	81.2	44	20	AA2016	Hepatitis C virus
38	121	81.2	44	21	AA2017	Human hepatitis C
39	121	81.2	45	21	AA2018	HCV core-envelope
40	121	81.2	55	13	AA2019	HCV core-envelope
41	121	81.2	55	13	AA2020	HCV core-envelope
42	121	81.2	55	13	AA2021	HCV core-envelope
43	121	81.2	55	13	AA2022	HCV core-envelope
44	121	81.2	55	13	AA2023	HCV core-envelope
45	121	81.2	57	13	AA2024	Non-A, Non-B Hepat

ALIGNMENTS

RESULT 1
AA1980
ID AAY06682 standard; Protein; 28 AA.
XX AAY06682:
AC
XX
XX 17-JUN-1999 (first entry)
DT
DE NC mosaic protein amino acid fragment J.
XX
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
KW restriction endonuclease assisted ligation; vaccination.
XX
XX Hepatitis C virus.
OS
XX
XX WO9910506-A1.
XX
XX 04-MAR-1999.
XX
XX 21-AUG-1998; 98WO-US17385.
XX
XX 25-AUG-1997; 97US-0921887.
PR
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Fields HA, Khudyakov YE;
PI
XX WPI; 1999-204671/17.
DR
XX New mosaic protein, comprising a plurality of homologous antigenic
PT peptides from different genotypes of a species - useful for
PT detecting hepatitis infection in an individual

XX Claim 5; Fig 9; 66pp; English.

XX The invention relates to a mosaic protein, comprising a plurality of

CC homologous antigenic peptides from different genotypes of a species. The

CC antigenic peptides are from nucleocapsid (NC) proteins. A method for

CC synthesising an artificial gene that encodes the mosaic protein is also

CC provided. The method is designated restriction endonuclease assisted

CC ligation (REAL). The mosaic protein and the artificial mosaic protein

CC are useful for detecting a hepatitis infection in an individual. The

CC mosaic gene and protein is also useful for vaccination against the

CC infection, especially hepatitis C. The method of synthesizing the

CC artificial gene and the resulting mosaic protein improve the sensitivity,

CC spectrum of immunoreactivity, and antigen specificity of enzyme

CC immunoassays. This provides improved detection of hepatitis C virus.

CC Sequences AAY06673-683 represent amino acid sequence of each monomer

CC comprising the NC mosaic protein.

XX Sequence 28 AA;

SQ Query Match 100.0%; Score 149; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 5.2e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGOIVG 28

DB 1 PKPKRQTKRNTLRPKNVKFPAGGOIVG 28

RESULT 2

AAR94462

ID AAR94462 standard; Protein; 3023 AA.

XX AAR94462;

AC AAR94462;

DT 20-SEP-1996 (first entry)

XX Hepatitis C virus polypeptide.

XX Hepatitis C virus; antibody: detection; diagnosis; vaccine;

KW classify; subtype.

XX Hepatitis C virus.

OS

XX

XX Location/Qualifiers

FT 1505..1520

FT Peptide

FT /note= "this part of the sequence is missing from

FT the specification"

FT Peptide

FT 2433..2448

FT /note= "this part of the sequence is missing from

FT the specification"

AA JP08056672-A.

PN

XX

XX

XX 05-MAR-1996.

XX

XX 26-AUG-1994; 94JP-0223933.

XX

XX 26-AUG-1994; 94JP-0223933.

XX

XX (SAYA/) SAYAMA K.

XX

XX WPI; 1996-182301/19.

DR N-PSDB; AAT13279.

XX

XX Hepatitis C virus genomic RNA, DNA and related proteins - useful for

PT detection, diagnosis and identification of hepatitis C virus

PT sub-type

XX

XX Claim 4; Page 16-23; 25pp; Japanese.

PS

XX The present sequence is a polypeptide comprising a 3023 amino acid

CC sequence which is encoded by a cDNA (AAT13279) to a hepatitis C virus

CC (HCV) genomic RNA. The protein can be easily detected by antibodies

CC in an assay for the detection of HCV. The DNA and the protein are

CC useful for classifying the subtype of HCV. At least a part of the

CC protein may be used as a vaccine against HCV.

XX Sequence 3023 AA;

SQ Query Match 100.0%; Score 149; DB 17; Length 3023;

Best Local Similarity 100.0%; Pred. No. 7e-12;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGOIVG 28

DB 5 PKPKRQTKRNTLRPKNVKFPAGGOIVG 32

RESULT 3

AAR34473

ID AAR34473 standard; Protein; 470 AA.

XX AAR34473;

AC AAR34473;

DT 30-JUL-1993 (first entry)

XX Encoded by Hepatitis C virus clone JK3-A.

XX HCV; non-A, non-B hepatitis virus; NANBH; liver disease;

KW polymerase chain reaction; diagnostic method.

XX Hepatitis C virus.

OS

XX JP05068562-A.

PN

XX

XX 23-MAR-1993.

XX

XX 30-MAY-1991; 91JP-0153736.

XX

XX 30-MAY-1991; 91JP-0153736.

XX (SANW) SANWA KAGAKU KENKYUSHO CO.

PA

XX

XX WPI; 1993-130638/16.

DR N-PSDB; AAQ40431.

XX

XX DNA and cDNA of hepatitis C virus - useful as probes for

PT diagnosing HCV infection

PT

XX Claim 4; Page 26-28; 44pp; Japanese.

PS

XX

XX cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B

CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR

CC amplification, including clone JK3-A. Primer/probes derived from the

CC sequences of these clones can be used in diagnostic assays for HCV.

CC See AAQ40425-Q40439.

XX

XX Sequence 470 AA;

SQ Query Match 87.2%; Score 130; DB 14; Length 470;

Best Local Similarity 82.1%; Pred. No. 4.8e-10;

Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGOIVG 28

DB 5 PKPKRQTKRNTLRPKNVKFPAGGOIVG 32

RESULT 4

AAR34474

ID AAR34474 standard; Protein; 470 AA.

XX AAR34474;

AC AAR34474;

XX

XX 30-JUL-1993 (first entry)

```

XX DE Encoded by Hepatitis C virus clone JK3-B.
XX
XX HCV; non-A, non-B hepatitis virus; NANBH; liver disease;
KW polymerase chain reaction; diagnostic method.
XX
XX Hepatitis C virus.
OS
XX JP05068562-A.
XX
XX 23-MAR-1993.
XX
XX 30-MAY-1991; 91JP-0153736.
XX
XX 30-MAY-1991; 91JP-0153736.
XX
XX (SANW ) SANWA KAGAKU KENKYUSHO CO.
PA
XX WPI; 1993-130638/16.
XX N-PSDB; AAO40432.
XX
XX DNA and cDNA of hepatitis C virus - useful as probes for
PT diagnosing HCV infection
XX
XX Claim 4; Page 28-30; 44pp; Japanese.
XX
XX cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B
CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR
CC amplification, including clone JK3-B. Primer/probes derived from the
CC sequences of these clones can be used in diagnostic assays for HCV.
CC See AAO40425-Q40439.
XX
XX Sequence 470 AA;
SQ
    Query Match      87.2%; Score 130; DB 14; Length 470;
    Best Local Similarity 82.1%; Pred. No. 4.8e-10;
    Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28
   |||:|||||:|||||:|||||
DB 5 PKPQRTKRTIRRPQDVKFPGGQIVG 32
   |||:|||||:|||||:|||||

RESULT 5
AAR34475
ID AAR34475 standard; Protein; 470 AA.
XX
XX AAR34475;
XX
XX 30-JUL-1993 (first entry)
XX
XX Encoded by Hepatitis C virus clone JK3-C.
XX
XX HCV; non-A, non-B hepatitis virus; NANBH; liver disease;
KW polymerase chain reaction; diagnostic method.
XX
XX Hepatitis C virus.
OS
XX JP05068562-A.
XX
XX 23-MAR-1993.
XX
XX 30-MAY-1991; 91JP-0153736.
XX
XX 30-MAY-1991; 91JP-0153736.
XX
XX (SANW ) SANWA KAGAKU KENKYUSHO CO.
PA
XX WPI; 1993-130638/16.
XX N-PSDB; AAO40433.
XX
XX DNA and cDNA of hepatitis C virus - useful as probes for
PT diagnosing HCV infection

```

```

XX Claim 4; Page 30-32; 44pp; Japanese.
XX
XX cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B
CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR
CC amplification, including clone JK3-C. Primer/probes derived from the
CC sequences of these clones can be used in diagnostic assays for HCV.
CC See AAO40425-Q40439.
XX
XX Sequence 470 AA;
SQ
    Query Match      87.2%; Score 130; DB 14; Length 470;
    Best Local Similarity 82.1%; Pred. No. 4.8e-10;
    Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28
   |||:|||||:|||||:|||||
DB 5 PKPQRTKRTIRRPQDVKFPGGQIVG 32
   |||:|||||:|||||:|||||

RESULT 6
AAY06673
ID AAY06673 standard; Protein; 28 AA.
XX
XX AAY06673;
XX
XX 17-JUN-1999 (first entry)
XX
XX NC mosaic protein amino acid fragment A.
DE
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
KW restriction endonuclease assisted ligation; vaccination.
XX
XX Hepatitis C virus.
OS
XX WO9910506-A1.
XX
XX 04-MAR-1999.
XX
XX 21-AUG-1998; 98WO-US17385.
XX
XX 25-AUG-1997; 97US-0921887.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Fields HA, Khudyakov YE;
XX
XX WPI; 1999-204671/17.
XX
XX New mosaic protein, comprising a plurality of homologous antigenic
PT peptides from different genotypes of a species - useful for
PT detecting hepatitis infection in an individual
XX
XX Claim 5; Fig 9; 66pp; English.
XX
XX The invention relates to a mosaic protein, comprising a plurality of
CC homologous antigenic peptides from different genotypes of a species. The
CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
CC synthesizing an artificial gene that encodes the mosaic protein is also
CC provided. The method is designated restriction endonuclease assisted
CC ligation (REAL). The mosaic protein and the artificial mosaic protein
CC are useful for detecting a hepatitis infection in an individual. The
CC mosaic gene and protein is also useful for vaccination against the
CC infection, especially hepatitis C. The method of synthesizing the
CC artificial gene and the resulting mosaic protein improve the sensitivity,
CC spectrum of immunoreactivity, and antigen specificity of enzyme
CC immunoassays. This provides improved detection of hepatitis C virus.
CC Sequences AAY06673-683 represent amino acid sequence of each monomer
CC comprising the NC mosaic protein.
XX
XX Sequence 28 AA;
SQ
    Query Match      84.6%; Score 126; DB 20; Length 28;

```

Best Local Similarity 78.6%; Pred. No. 9.1e-11;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
||||:|||||:|||||:|||||

Db 1 PKPQRTKRTNTRRPQDVKFPGGQIVG 28
||||:|||||:|||||:|||||

RESULT 7
AAB71258
ID AAB71258 standard; protein; 189 AA.
XX
AC AAB71258;
XX
DT 18-NOV-2002 (first entry)
XX
DE HCV type 3 capsid protein fragment.
XX
KW Capsid protein; attenuated vaccine; virucide; antiinflammatory;
KW hepatotropic; yellow fever; Japanese encephalitis; dengue;
KW classical swine fever; bovine viral diarrhoea; hepatitis C.
XX
Hepatitis C virus.
XX WO200266621-A1.
XX
XX 29-AUG-2002.
XX
XX 11-FEB-2002; 2002WO-AT00046.
XX
XX 21-FEB-2001; 2001AT-0000272.
XX
XX (HEINZ) HEINZ F X.
XX (MAND/) MANDL C.
XX
XX Heinz FX, Mandl C;
XX
XX WPI; 2002-667064/71.
XX
XX Attenuated flavivirus live vaccine, useful for protection against e.g.
XX yellow fever, comprises virus with attenuating deletion of amino acids
XX from the capsid protein.
XX
XX Disclosure; Fig 2; 30pp; German.
XX
XX This invention describes a novel attenuated flavivirus live vaccine
XX comprising a flavivirus mutant that has a deletion of at least 4
XX consecutive amino acids from the capsid protein, provided that the
XX C-terminal hydrophobic region is not affected by the deletion. The
XX vaccine of the invention has virucide, antiinflammatory and hepatotropic
XX activity. The attenuated vaccine, and similar nucleic acid vaccines that
XX encode the mutated capsid protein, are useful for protection against a
XX wide range of flavivirus diseases, e.g. yellow fever, Japanese
XX encephalitis, dengue, classical swine fever, bovine viral diarrhoea and
XX hepatitis C. The specified deletion: (i) produces a reliably attenuated
XX virus that does not revert to virulence; (ii) is exactly defined and does
XX not effect immune responses to important proteins; and (iii) can not
XX generate a non-natural virus by recombination. The mutant viruses
XX eliminate the need to produce large amounts of infectious/virulent
XX viruses, and can be produced with less expense. The protective response
XX to flavivirus lasts significantly longer than that to killed vaccines.
XX This sequence represents a fragment of the capsid protein from Hepatitis
XX C virus (HCV) type 1 described in the disclosure of the invention.
XX
XX Sequence 189 AA;

Query Match 84.6%; Score 126; DB 23; Length 189;
Best Local Similarity 78.6%; Pred. No. 6.8e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
||||:|||||:|||||:|||||

Db 4 PKPQRTKRTNTRRPQDVKFPGGQIVG 31
||||:|||||:|||||:|||||

us-09-491-146a-32.rag

Best Local Similarity 78.6%; Pred. No. 9.1e-11;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
||||:|||||:|||||:|||||

Db 1 PKPQRTKRTNTRRPQDVKFPGGQIVG 28
||||:|||||:|||||:|||||

RESULT 8
AAR92968
ID AAR92968 standard; Protein; 191 AA.
XX
AC AAR92968;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate HK10 core protein.
XX
KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN WO9605315-A2.
XX
PD 22-FEB-1996.
XX
XX 15-AUG-1995; 95WO-US10398.
XX
XX 15-AUG-1994; 94US-0290665.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (USSH) US SEC DEPT HEALTH.
XX
XX Bukh J, Miller RH, Purcell RH;
XX
XX WPI; 1996-139709/14.
XX N-PSDB; AAT16642.
XX
XX DNA and amino acid sequence of HCV envelope 1 and core proteins -
XX used to determine HCV genotype and as vaccines against HCV infection
XX
XX Claim 4; Page 207; 340pp; English.
XX
XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX isolates. Isolated cDNA sequences are used for the prodn. of primers
XX useful for detecting the presence of HCV in a sample, the primers
XX are also useful for HCV genotyping. Proteins encoded by the cDNAs
XX can be used in vaccines for immunising against HCV infection. The
XX proteins may also be used to detect antibodies against HCV in serum,
XX saliva, lymphocytes or other mononuclear cells. The antibodies may
XX be used in the prevention of HCV infection.
XX
XX Sequence 191 AA;

Query Match 84.6%; Score 126; DB 17; Length 191;
Best Local Similarity 78.6%; Pred. No. 6.8e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
||||:|||||:|||||:|||||

Db 5 PKPQRTKRTNTRRPQDVKFPGGQIVG 32
||||:|||||:|||||:|||||

RESULT 9
AAR92969
ID AAR92969 standard; Protein; 191 AA.
XX
AC AAR92969;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate S52 core protein.
XX
KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX

PN WO9613590-A2.
 XX 09-MAY-1996.
 XX 23-OCT-1995; 95WO-EP04155.
 XX 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX (INNO-) INNOGENETICS NV.
 XX Maertens G, Stuyver L;
 PI WPI; 1996-251460/25.
 DR N-PSDB; AAT27988.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub.type
 PT - used to develop probes and primers for new sub.types and vaccines
 PT to prevent and treat infection
 PT Claim 25; Fig 3; 150pp; English.
 The sequences AAR96526-R96578 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
 CC genome. This sequence represents amino acids 1-317 from the HCV types 7c
 CC and 8a isolates VN4.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 XX anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 XX Sequence 319 AA;
 SQ Query Match 84.6%; Score 126; DB 17; Length 319;
 Best Local Similarity 78.6%; Pred. No. 1.2e-09;
 Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 OY 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28
 DB 5 PKPQRTKRNTRRRPQDVKFPGGQIVG 32
 |||:|||||:|||||:|||||
 |||:|||||:|||||:|||||
 RESULT 12
 AAY06674
 ID AAY06674 standard; Protein; 28 AA.
 XX AAY06674;
 AC AAY06674;
 XX 17-JUN-1999 (first entry)
 DT NC mosaic protein amino acid fragment B.
 DE Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
 XX restriction endonuclease assisted ligation; vaccination.
 KW Hepatitis C virus.
 OS Hepatitis C virus.
 XX WO9910506-A1.
 PN 04-MAR-1999.
 XX 21-AUG-1998; 98WO-US17385.
 PF 25-AUG-1997; 97US-0921887.
 PR

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Fields HA, Khudyakov YE;
 XX WPI; 1999-204671/17.
 DR New mosaic protein, comprising a plurality of homologous antigenic
 XX peptides from different genotypes of a species - useful for
 PT detecting hepatitis infection in an individual
 PT Claim 5; Fig 9; 66pp; English.
 XX The invention relates to a mosaic protein, comprising a plurality of
 CC homologous antigenic peptides from different genotypes of a species. The
 CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
 CC synthesising an artificial gene that encodes the mosaic protein is also
 CC provided. The method is designated restriction endonuclease assisted
 CC ligation (REAL). The mosaic protein and the artificial mosaic protein
 CC are useful for detecting a hepatitis infection in an individual. The
 CC mosaic gene and protein is also useful for vaccination against the
 CC infection, especially hepatitis C. The method of synthesizing the
 CC artificial gene and the resulting mosaic protein improve the sensitivity,
 CC spectrum of immunoreactivity, and antigen specificity of enzyme
 CC immunoassays. This provides improved detection of hepatitis C virus.
 CC Sequences AAY06673-683 represent amino acid sequence of each monomer
 CC comprising the NC mosaic protein.
 XX Sequence 28 AA;
 SQ Query Match 83.9%; Score 125; DB 20; Length 28;
 Best Local Similarity 82.1%; Pred. No. 1.3e-10;
 Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 OY 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28
 DB 1 PKPQRTKRNTRRRPQDVKFPGGQIVG 28
 |||:|||||:|||||:|||||
 |||:|||||:|||||:|||||
 RESULT 13
 AAR96577
 ID AAR96577 standard; peptide; 74 AA.
 XX AAR96577;
 AC AAR96577;
 XX 14-MAR-1997 (first entry)
 DT Hepatitis C virus type 11a isolate FR19 amino acids 1-74.
 DE Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 XX PCR; primer; probe; antibody; infection.
 KW Hepatitis C virus.
 OS Hepatitis C virus.
 XX WO9613590-A2.
 PN 09-MAY-1996.
 XX 23-OCT-1995; 95WO-EP04155.
 XX 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX (INNO-) INNOGENETICS NV.
 XX Maertens G, Stuyver L;
 PI WPI; 1996-251460/25.
 DR N-PSDB; AAT27988.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub.type
 PT - used to develop probes and primers for new sub.types and vaccines
 PT to prevent and treat infection

XX PS Claim 25; Fig 3; 150pp; English.

CC The sequences AAR96526-R96578 represent novel sequences isolated from

CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,

CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,

CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'

CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the

CC genome. This sequence represents amino acids 1-74 from the HCV type 11a

CC isolate FR19.

CC The new HCV types were isolated from patients with chronic HCV from the

CC Benelux countries, France, Cameroon and Vietnam, because of their

CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR

CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were

CC sequenced either directly or partially and used to classify the new

CC viruses into (sub)types based on comparison with known sequences.

CC The sequences can be used to generate the peptides AAR96424-R96524. The

CC sequences can also be used to synthesise probes and primers for the

CC detection of HCV in a sample. The polypeptides can be used to detect

CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.

SV Sequence 74 AA;

Query Match 83.9%; Score 125; DB 17; Length 74;

Best Local Similarity 82.1%; Pred. No. 3.5e-10;

Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28

Db 5 PKPQRQTKRNTLRPKNVKFPAGGQIVG 32

RESULT 14

AAR92943

ID AAR92943 standard; Protein; 191 AA.

AC AAR92943;

XX 02-OCT-1996 (first entry)

DT Hepatitis C virus isolate S45 core protein.

DE HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

KX Hepatitis C virus.

OS WO9605315-A2.

XX 22-FEB-1996.

PF 15-AUG-1995; 95WO-US10398.

XX 15-AUG-1994; 94US-0290665.

PR (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (USSH) US SEC DEPT HEALTH.

XX Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

DR N-PSDB; AAT16617.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

PS Claim 4; Page 186-187; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers

CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX Sequence 191 AA;

Query Match 83.9%; Score 125; DB 17; Length 191;

Best Local Similarity 75.0%; Pred. No. 9.5e-10;

Matches 21; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28

Db 5 PKPQRQTKRNTLRPKNVKFPAGGQIVG 32

Search completed: August 7, 2003, 11:14:10

Job time : 39.6364 secs

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX Sequence 191 AA;

Query Match 83.9%; Score 125; DB 17; Length 191;

Best Local Similarity 82.1%; Pred. No. 9.5e-10;

Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28

Db 5 PKPQRQTKRNTLRPKNVKFPAGGQIVG 32

RESULT 15

AAR92970

ID AAR92970 standard; Protein; 191 AA.

AC AAR92970;

XX 02-OCT-1996 (first entry)

DT Hepatitis C virus isolate S2 core protein.

DE HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

KX Hepatitis C virus.

OS WO9605315-A2.

XX 22-FEB-1996.

PF 15-AUG-1995; 95WO-US10398.

XX 15-AUG-1994; 94US-0290665.

PR (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (USSH) US SEC DEPT HEALTH.

XX Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

DR N-PSDB; AAT16644.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

PS Claim 4; Page 209; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers

CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX Sequence 191 AA;

Query Match 83.9%; Score 125; DB 17; Length 191;

Best Local Similarity 75.0%; Pred. No. 9.5e-10;

Matches 21; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28

Db 5 PKPQRQTKRNTLRPKNVKFPAGGQIVG 32

Search completed: August 7, 2003, 11:14:10

Job time : 39.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-32

Perfect score: 149

Sequence: 1 PKPKRQTKRNTLRPNKVNKFPAGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

al number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	87.2	782	2 S19875	genome polyprotein
2	129	86.6	411	2 PC2060	genome polyprotein
3	126	84.6	114	2 S41359	genome polyprotein
4	126	84.6	114	2 S41358	genome polyprotein
5	126	84.6	124	2 S41360	genome polyprotein
6	126	84.6	411	2 PC2061	genome polyprotein
7	126	84.6	492	2 S41288	genome polyprotein
8	125	83.9	115	2 S41351	genome polyprotein
9	125	83.9	115	2 S41349	genome polyprotein
10	125	83.9	640	2 JQ1584	genome polyprotein
11	123	82.6	108	2 S41356	genome polyprotein
12	123	82.6	3010	1 A45573	genome polyprotein
13	122	81.9	88	2 S21336	genome polyprotein
14	122	81.9	123	2 S41361	genome polyprotein
15	121	81.2	108	2 S41353	genome polyprotein
16	121	81.2	108	2 S41355	genome polyprotein
17	121	81.2	108	2 S41357	genome polyprotein
18	121	81.2	108	2 S41348	genome polyprotein
19	121	81.2	112	2 S41371	genome polyprotein
20	121	81.2	112	2 S41341	genome polyprotein
21	121	81.2	114	2 S41370	genome polyprotein
22	121	81.2	114	2 S41369	genome polyprotein
23	121	81.2	114	2 S41368	genome polyprotein
24	121	81.2	115	2 S41342	genome polyprotein
25	121	81.2	115	2 S41344	genome polyprotein
26	121	81.2	115	2 S41350	genome polyprotein
27	121	81.2	115	2 S41354	genome polyprotein
28	121	81.2	115	2 S41345	genome polyprotein
29	121	81.2	115	2 S41347	genome polyprotein

30 121 81.2 115 2 S41343 genome polyprotein
31 121 81.2 118 2 S41346 genome polyprotein
32 121 81.2 369 2 S21471 genome polyprotein
33 121 81.2 441 2 S12707 genome polyprotein
34 121 81.2 513 2 PC1284 genome polyprotein
35 121 81.2 520 2 JQ1925 polyprotein - hepa
36 121 81.2 523 2 JQ1926 polyprotein - hepa
37 121 81.2 550 2 JH0711 genome polyprotein
38 121 81.2 782 2 S19876 genome polyprotein
39 121 81.2 782 2 S18031 genome polyprotein
40 121 81.2 782 2 S18032 genome polyprotein
41 121 81.2 787 2 P05677 hypothetical prote
42 121 81.2 874 2 JQ0883 genome polyprotein
43 121 81.2 874 2 JQ0881 genome polyprotein
44 121 81.2 876 2 PC2219 polypeptide - hepa
45 121 81.2 3010 1 GNVVTC genome polyprotein

ALIGNMENTS

RESULT 1

S19875
genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 prot
C:Species: hepatitis C virus
A:Variety: isolate JK3
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S19875
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Vir
A:Reference number: S18029
A:Accession: S19875
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61592; NID:G59482; PIDN:CAA43789.1; PID:G59483
A:Experimental source: isolate JK3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstruc
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 87.2%; Score 130; DB 2; Length 782;

Best Local Similarity 82.1%; Pred. No. 4.1e-10;

Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPNKVNKFPAGGQIVG 28

Db 5 PKPQRTKRTNTRRQDVAKFPGGQIVG 32

RESULT 2

PC2060
genome polyprotein N1 - hepatitis C virus
N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2060
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France
A:Reference number: PC2060; MUID:94197744; PMID:8147893
A:Accession: PC2060
A:Molecule type: mRNA
A:Residues: 1-411 <LIJ>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural pro
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match      86.6%; Score 129; DB 2; Length 411;
Best Local Similarity 78.6%; Pred. No. 3.1e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRRPKNKVKFPAGGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRTKRTNTRRPQDVKFPGGGQIVG 32

RESULT 3
S41359
genome polyprotein - hepatitis C virus (genotype 3, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41359
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41359
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229462; NID:g443886; PIDN:CAA82600.1; PID:g443887
A:Experimental source: genotype 3, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <NAT>

Query Match      84.6%; Score 126; DB 2; Length 114;
Best Local Similarity 78.6%; Pred. No. 2.3e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRRPKNKVKFPAGGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRTKRTNTRRPQDVKFPGGGQIVG 32

RESULT 4
S41358
genome polyprotein - hepatitis C virus (genotype 3, N1) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41358
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41358
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:229461; NID:g443884; PIDN:CAA82599.1; PID:g443885
A:Experimental source: genotype 3, N1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <NAT>

Query Match      84.6%; Score 126; DB 2; Length 114;
Best Local Similarity 78.6%; Pred. No. 2.3e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRRPKNKVKFPAGGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRTKRTNTRRPQDVKFPGGGQIVG 32

RESULT 5
S41360
genome polyprotein - hepatitis C virus (genotype 3, N3) (fragment)
```

```
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41360
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41360
A:Molecule type: genomic RNA
A:Residues: 1-124 <VAN>
A:Cross-references: EMBL:229463; NID:g443888; PIDN:CAA82601.1; PID:g443889
A:Experimental source: genotype 3, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-124/Product: core protein #status predicted <NAT>

Query Match      84.6%; Score 126; DB 2; Length 124;
Best Local Similarity 78.6%; Pred. No. 2.5e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRRPKNKVKFPAGGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRTKRTNTRRPQDVKFPGGGQIVG 32

RESULT 6
PC2061
genome polyprotein N2 - hepatitis C virus
N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2061
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France.
A:Reference number: PC2060; MUID:94197744; PMID:8147893
A:Accession: PC2061
A:Molecule type: mRNA
A:Residues: 1-411 <LIJ>
A:Cross-references: GB:LI2355; NID:g410169; PIDN:AA20155.1; PID:g410170
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural prot
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      84.6%; Score 126; DB 2; Length 411;
Best Local Similarity 78.6%; Pred. No. 8e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRRPKNKVKFPAGGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRTKRTNTRRPQDVKFPGGGQIVG 32

RESULT 7
S41288
genome polyprotein - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein; NS1 protein
C:Species: hepatitis C virus
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41288
R:Seelig, R.
submitted to the EMBL Data Library, December 1993
A:Reference number: S41288
A:Accession: S41288
A:Molecule type: genomic RNA
A:Residues: 1-492 <SEE>
A:Cross-references: EMBL:X76918
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; nonstructural protein;
```

F:1-191/Product: core protein #status predicted <COR>
 F:192-372/Product: envelope protein #status predicted <ENV>
 F:373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match 84.6%; Score 126; DB 2; Length 492;
 Best Local Similarity 78.6%; Pred. No. 9.5e-10;
 Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPNKPKVPPAGGQIVG 28
 |||:|||||:|||||:|||||
 Db 5 PKPQRQTKRNTLRPNKPKVPPAGGQIVG 32

RESULT 8

S41351
 genome polyprotein - hepatitis C virus (genotype 1, N9) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N9
 Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 Accession: S41351

van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341

A:Accession: S41351
 A:Molecule type: genomic RNA
 A:Residues: 1-115 <VAN>
 A:Cross-references: EMBL:Z29454
 A:Experimental source: genotype 1, N9
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-115/Product: core protein #status predicted <MAT>

Query Match 83.9%; Score 125; DB 2; Length 115;
 Best Local Similarity 82.1%; Pred. No. 3.2e-10;
 Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPNKPKVPPAGGQIVG 28
 |||:|||||:|||||:|||||
 Db 5 PKPQRQTKRNTLRPNKPKVPPAGGQIVG 32

RESULT 9

S41349
 genome polyprotein - hepatitis C virus (genotype 1, N7) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N7
 Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 Accession: S41349

van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41349
 A:Molecule type: genomic RNA
 A:Residues: 1-115 <VAN>
 A:Cross-references: EMBL:Z29452; NID:g443866; PIDN:CAA82590.1; PID:g443867
 A:Experimental source: genotype 1, N7
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-115/Product: core protein #status predicted <MAT>

Query Match 83.9%; Score 125; DB 2; Length 115;
 Best Local Similarity 82.1%; Pred. No. 3.2e-10;
 Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPNKPKVPPAGGQIVG 28
 |||:|||||:|||||:|||||
 Db 5 PKPQRQTKRNTLRPNKPKVPPAGGQIVG 32

RESULT 10

JQ1584
 genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
 N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural
 C:Species: hepatitis C virus
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
 C:Accession: JQ1584
 R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
 J. Gen. Virol. 73, 1521-1525, 1992
 A:Title: Cloning and sequencing of the structural region and expression of putative
 A:Reference number: JQ1584; MUID:92300349; PMID:1313944

A:Accession: JQ1584
 A:Molecule type: genomic RNA
 A:Residues: 1-640 <KUM>
 A:Cross-references: GB:X84079; NID:g643119; PIDN:CAA58888.1; PID:g643120
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; p
 F:1-191/Product: core protein C #status predicted <CPC>
 F:192-389/Product: envelope protein E1 #status predicted <EE1>
 F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predi
 F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn)

Query Match 83.9%; Score 125; DB 2; Length 640;
 Best Local Similarity 82.1%; Pred. No. 1.7e-09;
 Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPNKPKVPPAGGQIVG 28
 |||:|||||:|||||:|||||
 Db 5 PKPQRQTKRNTLRPNKPKVPPAGGQIVG 32

RESULT 11

S41356
 genome polyprotein - hepatitis C virus (genotype 2, N5) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N5
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41356
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341

A:Accession: S41356
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:Z29459
 A:Experimental source: genotype 2, N5
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <MAT>

Query Match 82.6%; Score 123; DB 2; Length 108;
 Best Local Similarity 78.6%; Pred. No. 5.7e-10;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPNKPKVPPAGGQIVG 28
 |||:|||||:|||||:|||||
 Db 5 PKPQRQTKRNTLRPNKPKVPPAGGQIVG 32

RESULT 12

A45573
 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nons
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hiji
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carr
 A:Reference number: A45573; MUID:92295714; PMID:1318627

A:Accession: A45573
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <VAN>
A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BA01943.1; PID:g221613

A:Experimental source: HCV-JT
A>Note: Sequence extracted from NCBI backbone (NCBI:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine protease; capsid protein C #status predicted <CPC>
F:2-115/Product: capsid protein C #status predicted <EPM>
F:192-389/Product: major envelope protein M #status predicted <EPM>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 82.6%; Score 123; DB 1; Length 3010;
Best Local Similarity 78.6%; Pred. No. 1.4e-08;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGGQIVG 28
||||:||||| |||:||||| |||
Db 5 PKPQRTKNTYRRPDQVRFPGGGQIVG 32

RESULT 13

S21336
genome polyprotein S29 (core protein region) - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
C:Accession: S21336
R:Sato, A.

submitted to the EMBL Data Library, April 1992

A:Description: A sensitive serodiagnosis of hepatitis C virus infection with two cloned

A:Reference number: S21336

A:Accession: S21336

A:Molecule type: genomic RNA

A:Residues: 1-88 <SAT>

A:Cross-references: EMBL:X65548; NID:g59492; PIDN:CAA46517.1; PID:g59493

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: polyprotein

Query Match 81.9%; Score 122; DB 2; Length 88;
Best Local Similarity 78.6%; Pred. No. 6.5e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGGQIVG 28
||||:||||| |||:||||| |||
Db 11 PKPQRTKNTSRRPQDVRFPGGGQIVG 38

RESULT 14

S41361

genome polyprotein - hepatitis C virus (genotype 3, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 3, N4

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41361

R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41361

A:Molecule type: genomic RNA

A:Residues: 1-123 <VAN>

A:Cross-references: EMBL:Z29464; NID:g443890; PIDN:CAA82602.1; PID:g443891

A:Experimental source: genotype 3, N4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-123/Product: core protein #status predicted <MAT>

Query Match 81.9%; Score 122; DB 2; Length 123;
Best Local Similarity 75.0%; Pred. No. 9e-10;
Matches 21; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGGQIVG 28
||||:||||| |||:||||| |||
Db 5 PKPQRTKNTYRRPDQVRFPGGGQIVG 32

RESULT 15

S41353

genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 2, N2

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41353

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41353

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:Z29456

A:Experimental source: genotype 2, N2

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 81.2%; Score 121; DB 2; Length 108;
Best Local Similarity 78.6%; Pred. No. 1.1e-09;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGGQIVG 28
||||:||||| |||:||||| |||
Db 5 PKPQRTKNTYRRPDQVRFPGGGQIVG 32

Search completed: August 7, 2003, 11:21:49
Job time: 9.54545 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146A-32

Perfect score: 149

Sequence: 1 PKPKRQTKRNTLRPNKVPKPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

al number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123	82.6	3010	1	POLG_HCVJT
2	121	81.2	513	1	POLG_HCVJ2
3	121	81.2	520	1	POLG_HCVH4
4	121	81.2	520	1	POLG_HCVHK
5	121	81.2	737	1	POLG_HCVJ5
6	121	81.2	737	1	POLG_HCVJ7
7	121	81.2	3010	1	POLG_HCVBK
8	121	81.2	3010	1	POLG_HCVJA
9	121	81.2	3011	1	POLG_HCVH
10	121	81.2	3033	1	POLG_HCVJ6
11	121	81.2	3033	1	POLG_HCVJ8
12	114	76.5	3010	1	POLG_HCVTW
13	113	75.8	3011	1	POLG_HCV1
14	52	34.9	332	1	PI11_MOUSE
15	51.5	34.6	3988	1	POLG_BVDVN
16	50.5	33.9	202	1	HL_LYCPN
17	50	33.6	794	1	FURI_HUMAN
18	49	32.9	4725	1	DYHC_DICDI
19	47.5	31.9	218	1	H12_RAT
20	47	31.5	58	1	RL29_YEAST
21	47	31.5	237	1	SGS_DROME
22	47	31.5	477	1	DXR_ARATH
23	47	31.5	1960	1	TF20_HUMAN
24	47	31.5	1983	1	TF20_MOUSE
25	46.5	31.2	159	1	RL29_MOUSE
26	46.5	31.2	584	1	VATA_METH
27	46.5	31.2	906	1	AD22_HUMAN
28	46.5	31.2	1236	1	POLS_WEEV
29	46	30.9	240	1	H12_VOLCA
30	46	30.9	517	1	LADI_HUMAN
31	45.5	30.5	3898	1	POLG_BVDVS
32	45	30.2	101	1	RL5_RABIT
33	45	30.2	152	1	DTD_THEAQ

34 45 30.2 194 1 H1_SALTR
35 45 30.2 296 1 RL5_MOUSE
36 45 30.2 296 1 RL5_RAT
37 45 30.2 469 1 APTD_DROME
38 45 30.2 793 1 FURI_MOUSE
39 44.5 29.9 303 1 YBL_XENLA
40 44.5 29.9 1679 1 YMF9_YEAST
41 44 29.5 345 1 PPAN_ARATH
42 44 29.5 550 1 COCH_HUMAN
43 44 29.5 590 1 SYKC_YEAST
44 44 29.5 613 1 CGAA_CLOBI
45 44 29.5 797 1 FURI_BOVIN

ALIGNMENTS

RESULT 1

POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RA "Molecular cloning of hepatitis C virus genome from a single Japanese
carrier: sequence variation within the same individual and among
infected individuals.";
RL Virus Res. #3:39-53(1992).
CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -|- HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC {RNA}(N).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D11168; BAA01943.1; -
CC PIR; A45573; A45573.
CC DR PDB; 1A1Q; 25-MAR-98.
CC DR PDB; 1JXP; 14-JAN-98.
CC DR MEROPS; S29.001; -
CC DR MEROPS; U39.001; -
CC InterPro; IPR001410; DEAD.

DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR004190; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RGR.
 DR InterPro: IPR007095; RNA_pol_DS_Ps.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RGRP; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS3 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS4A (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 1863 2013 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 2014 3010 POTENTIAL.
 FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 ATP (POTENTIAL).
 FT NP_BIND 1230 1237 DECH BOX.
 FT SITE 1316 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C7435D642BB CRC64;
 Query Match 82.6%; Score 123; DB 1; Length 3010;
 Best Local Similarity 78.6%; Pred. No. 1e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 5 PKPQKTKRNTYRRPQDVKFPGGQIVG 32
 RESULT 2
 POLG_HCVJ2 STANDARD; PRT; 513 AA.
 ID P27959;
 AC P27959;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HC-J2) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT *Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.*;
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 EMBL: D10074; BAA00968.1; .
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 KW polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 >513 MAJOR ENVELOPE PROTEIN NS1 (POTENTIAL).
 FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 513

SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 81.2%; Score 121; DB 1; Length 513;
Best Local Similarity 78.6%; Pred. No. 3.2e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFFPAGGQIVG 28
||||:||||| |||:||||| |||||
DB 5 PKPQRTKRNTRRPQDVKFFGGQIVG 32

RESULT 3

ID POLG_HCVH4 STANDARD; PRT; 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
(GP68) (GP70) (NS1)] (Fragment).
DE Hepatitis C virus (isolate HCV-476) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
isolated from a patient who contracted hepatitis during an epidemic
of non-A, non-B hepatitis in Japan."
RL J. Gen. Virol. 73:2725-2729(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D10688; BAA01530.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM BY SIMILARITY.
FT CARBOHYD 196 196
FT CARBOHYD 209 196
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM BY SIMILARITY.
FT CARBOHYD 196 196
FT CARBOHYD 209 196
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449

FT NON_TER 520 520

SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 81.2%; Score 121; DB 1; Length 520;
Best Local Similarity 78.6%; Pred. No. 3.3e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFFPAGGQIVG 28
||||:||||| |||:||||| |||||
DB 5 PKPQRTKRNTRRPQDVKFFGGQIVG 32

RESULT 4

ID POLG_HCVHK STANDARD; PRT; 520 AA.
AC Q01403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
(GP68) (GP70) (NS1)] (Fragment).
DE Hepatitis C virus (isolate HCV-KF) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
isolated from a patient who contracted hepatitis during an epidemic
of non-A, non-B hepatitis in Japan."
RL J. Gen. Virol. 73:2725-2729(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D10687; BAA01529.1; -
DR PIR; JQ1925; JQ1925.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM BY SIMILARITY.
FT CARBOHYD 196 196
FT CARBOHYD 209 196
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM BY SIMILARITY.
FT CARBOHYD 196 196
FT CARBOHYD 209 196
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418

```

FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match
Best Local Similarity 81.2%; Score 121; DB 1; Length 520;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPRKPKVFPAGGQIVG 28
    |||:||||| |||:||||| |||
Db 5 PKPQKTKRNTNRPRQDVKFGGQIVG 32

RESULT 5
POLG_HCVJ5 STANDARD; PRT; 737 AA.
ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1 and NS2] (Fragment).
DE Hepatitis C virus (isolate HC-J5) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
homology to reported isolates: comparative study of four distinct
genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
NS3 AND NS5 MAY PLAY A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; D10075; BAA00969.1;
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF0186062; HCV_NS1; 1.
CC Pfam; PF0186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

```

FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 737 AA; 81207 MW; 3AF69D82AD501B1 CRC64;

Query Match
Best Local Similarity 81.2%; Score 121; DB 1; Length 737;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPRKPKVFPAGGQIVG 28
    |||:||||| |||:||||| |||
Db 5 PKPQKTKRNTNRPRQDVKFGGQIVG 32

RESULT 6
POLG_HCVJ7 STANDARD; PRT; 737 AA.
ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
01-AUG-1992 (Rel. 23, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Genome polyprotein [contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1 and NS2] (Fragment).
DE Hepatitis C virus (isolate HC-J7) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
homology to reported isolates: comparative study of four distinct
genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
NS3 AND NS5 MAY PLAY A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; D10077; BAA00971.1;
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF0186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
 FT Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122P2 CRC64;

Query Match 81.28; Score 121; DB 1; Length 737;
 Best Local Similarity 78.68; Pred. No. 4.7e-10;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKROTNRNTLRPNKVPAGGQIVG 28
 |||:|:|||||:|||||:|||||
 DB 5 PKPQRTKRNTRRPQDVPFGGQIVG 32

RESULT 7
 POLG_HCVBK
 ID POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 NM 15-SEP-2003 (Rel. 42, Last annotation update)
 Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RN MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RN SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetevy L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits

RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moomaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M58335; AAA72945.1; -
 DR PIR; A38465; GNWVTC
 DR PDB; 1AJO: 25-MAR-98.
 DR PDB; 1JXP: 14-JAN-98.
 DR PDB; 1NS3: 08-APR-98.
 DR PDB; 1C2P: 15-NOV-00.
 DR PDB; 1CSJ: 08-NOV-99.
 DR PDB; 1GX5: 09-APR-02.
 DR PDB; 1GX6: 10-APR-02.
 DR PDB; 1QOV: 26-JUN-00.
 DR PDB; 8OHM: 20-APR-99.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RDRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN E (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1039 1047 STRAND.
 FT STRAND 1050 1050 STRAND.
 FT STRAND 1059 1063 STRAND.
 FT TURN 1068 1074 TURN.
 FT STRAND 1075 1076 STRAND.
 FT HELIX 1077 1081 HELIX.
 FT TURN 1082 1085 TURN.
 FT STRAND 1086 1087 STRAND.
 FT TURN 1090 1092 TURN.
 FT STRAND 1093 1094 STRAND.
 FT STRAND 1095 1097 STRAND.
 FT STRAND 1101 1103 STRAND.
 FT TURN 1104 1107 TURN.
 FT STRAND 1108 1112 STRAND.
 FT STRAND 1120 1120 STRAND.
 FT STRAND 1122 1122 STRAND.
 FT TURN 1129 1133 TURN.
 FT STRAND 1135 1136 STRAND.
 FT STRAND 1139 1144 STRAND.
 FT STRAND 1149 1157 STRAND.
 FT HELIX 1158 1161 HELIX.
 FT TURN 1162 1163 TURN.
 FT TURN 1165 1166 TURN.
 FT STRAND 1168 1171 STRAND.
 FT TURN 1172 1174 TURN.
 FT STRAND 1175 1186 STRAND.

FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFFD9C CRC64;
 Query Match 81.2%; Score 121; DB 1; Length 3010;
 Best Local Similarity 78.6%; Pred. No. 2e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 OY 1 PKPKRQTRNTLRPKNKVFPAGGQIVG 28
 DB 5 PKPQRKTRNTNRPPQDVKEFGGQIVG 32
 RESULT 8
 POLG_HCVJA
 ID POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (Isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9108550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 {RNA}(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90208; BAA14233.1; -;
 PIR; A39253; GNMVCG.

DR H5SP: P26663; LJXP.
 DR MEROPS: S29.001; --
 DR MEROPS: U39.001; --
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002168; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polypeptide: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1087
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 FT SSEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 81.2%; Score 121; DB 1; Length 3010;
 Best Local Similarity 78.6%; Pred. No. 2e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKNTLRPPKNVKEPAGQIVG 28
 |||:|:||||| ||||| |||||
 Db 5 PKPQRKTKNTNRPPQDVKKFPGGQIVG 32
 RESULT 9
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4a (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9205256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RA "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RA "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morganstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RA "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 of unwinding.";
 RL Structure 6:89-100(1998).
 CC -|- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -|- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -|- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -|- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -|- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -|- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -|- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -|- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

P26660;
 01-AUG-1992 (Rel. 23, Created)
 01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11113;
 RN [1]
 SEQUENCE FROM N.A.
 MEDLINE=92044440; PubMed=1658196;
 Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.,
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions".
 J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D00944; BAA00792.1; -
 DR PIR: J01303; J01303.
 DR HSSP: P27958; 1HEI.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_Psvir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.

DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 DR Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT INT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT CHAIN 3033 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 81.2%; Score 121; DB 1; Length 3033;
 Best Local Similarity 78.6%; Pred. No. 2.1e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 PKPKRQTKRNTLRPKMKVFPAGQIVG 28
 Db 5 PKPQKTKRNTNRPPQDVKFPGGQIVG 32
 ||||:||||| ||||:||||| |||||
 RESULT 11
 ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11115;
 RN [1]

SEQUENCE FROM N.A.
 MEDLINE:92230232; PubMed-1314459;
 Okamoto S., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 Fukuda S., Tsuda F., Mishiro S.;
 *Full-length sequence of a hepatitis C virus genome having poor
 homology to reported isolates: comparative study of four distinct
 genotypes.*;
 Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; D10988; BAA01761.1; -;
 CC PIR; A40250; GNMVJ8.
 CC HSSP; P27958; 1HEI.
 CC MEROPS; S29.001; -;
 CC MEROPS; U39.001; -;
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RdRp.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.
 CC Pfam; PF01539; HCV_env; 1.
 CC Pfam; PF01560; HCV_NS1; 1.
 CC Pfam; PF01538; HCV_NS2; 1.
 CC Pfam; PF02907; HCV_NS3; 1.
 CC Pfam; PF01006; HCV_NS4a; 1.
 CC Pfam; PF01001; HCV_NS4b; 1.
 CC Pfam; PF01506; HCV_NS5a; 1.
 CC Pfam; PF00998; Viral_RdRp; 1.
 CC ProDom; PD186062; HCV_NS1; 1.
 CC SMART; SM00487; DEXdc; 1.
 CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033

CHAIN 3033
 TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
 Query Match 81.2%; Score 121; DB 1; Length 3033;
 Best Local Similarity 78.6%; Pred. No. 2,1e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 OY 1 PKPKRQTKRNTLRPKNTKFPAGGQIVG 28
 DB 5 PKPQRKTKRNTNRPRQDVKFPGGQIVG 32
 RESULT 12
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 ID POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230232; PubMed-1314449;
 RA Chen P.-J., Lin M.H., Tai K.F., Liu P.C., Lin C.-J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M84754; -; NOT_ANNOTATED_CDS.
 DR PIR; A0244; GNKWTW.
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.

DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_Ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 vq 3D-structure.

FT CHAIN 1 115 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 116 191 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 192 383 CORE PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 POTENTIAL.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 76.5%; Score 114; DB 1; Length 3010;
 Best Local Similarity 77.8%; Pred. No. 2.2e-08;
 Matches 21; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 KPKRQTKRNTLRKPKNVKFPAGGQIVG 28
 ||:||||| |||:||||| |||||
 DB 6 KPQRKTKRNTNRFPQDVAFPGGQIVG 32

RESULT 13

POLG_HCV1 STANDARD; PRT; 3011 AA.
 ID POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV)
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 ON NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic Organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M62321; AAA45676.1; -.

DR PIR: A39166; GNWVC3.
DR PDB: 1AIV; 16-FEB-99.
DR MEROPS: S29.001; 25-NOV-98.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RGRP.
DR InterPro: IPR001650; Helicase.C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02507; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase.C; 1.
DR Pfam: PF00998; Viral_RGRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 224 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQL SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
Query Match 75.8%; Score 113; DB 1; Length 3011;
Best Local Similarity 71.4%; Pred. No. 3; le-08;
Matches 20; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 1 PKPKQTRKRLRRPKNVKFPAGGOIVG 28
DB 5 PKPQKKRNTNRRPDQVKFPGGOIVG 32
RESULT 14
PILL_MOUSE
ID PILL_MOUSE STANDARD; PRT; 332 AA.
AC Q9CZK5; Q91WZ9; Q9DOC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pin2-interacting protein X1 (TRF1-interacting protein 1) (Liver-
DE related putative tumor suppressor) (LPTS1) (67-11-3 protein).
GN PINX1 OR LPTS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21558139; PubMed=11701125;
RA Zhou X.Z.; Lu K.P.;
RT "The Pin2/TRF1-interacting protein PinX1 is a potent telomerase
RT inhibitor.";
RL Cell 107:347-359(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR.
RA Liao C.; Zhao M.; Li T.;
RT "The expression of mouse LPTS1, a homolog of human tumor suppressor
RT LPTS, in mouse liver.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J.; Shinagawa A.; Shibata K.; Yoshino M.; Itoh M.; Ishii Y.;
RA Arakawa T.; Hara A.; Fukunishi Y.; Konno H.; Adachi J.; Fukuda S.;
RA Aizawa K.; Izawa M.; Nishi K.; Kiyosawa H.; Kondo S.; Yamana K.I.;
RA Saïto T.; Okazaki Y.; Gojobori T.; Bono H.; Kasukawa T.; Saïto R.;
RA Kadota K.; Matsuda H.A.; Ashburner M.; Batalov S.; Casavant T.;
RA Fleischmann W.; Gaasterland T.; Gissi C.; King B.; Kochiwa H.;
RA Kuehl P.; Lewis S.; Matsuo Y.; Nikaido I.; Pesole G.; Quackenbush J.;
RA Schriml L.M.; Staib F.; Suzuki R.; Tomita M.; Wagner L.; Washio T.;
RA Sakai K.; Okido T.; Furuno M.; Aono H.; Baldarelli R.; Barsh G.;
RA Blake J.; Boffelli D.; Bojunga N.; Carninci P.; de Bonaldo M.F.;
RA Brownstein M.J.; Bult C.; Fletcher C.; Fujita M.; Gariboldi M.;
RA Gustincich S.; Hall D.; Hofmann M.; Hume D.A.; Kamiya M.; Lee N.H.;
RA Lyons P.; Marchionni L.; Mashima J.; Marzari J.; Mombert P.;
RA Nordone P.; Ring B.; Ringwald M.; Rodriguez I.; Sakamoto N.;
RA Sasaki H.; Sato K.; Schoenbach C.; Seva T.; Shibata Y.; Storch K.-F.;
RA Suzuki H.; Toyooka K.; Wang K.H.; Weitz C.; Wittaker C.; Wilming L.;
RA Wynshaw-Boris A.; Yoshida K.; Hasegawa Y.; Kawaji H.; Kohsaki S.;
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE OF 38-332 FROM N.A.
RC TISSUE=Embryo;
RA Schmidt T.;
RL Thesis (2001), University of Goettingen, Germany.
CC -!- FUNCTION: Inhibits telomerase activity. May inhibit cell
CC proliferation and act as tumor suppressor (By similarity).
CC -!- SUBUNIT: Binds to TERT and TRF1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear; in nucleoli and at telomere
CC speckles (By similarity).

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-32
Perfect score: 149
Sequence: 1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mbc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149	100.0	3023	12	Q81487 hepatitis c
2	137	91.9	419	12	Q81345 hepatitis c
3	132	88.6	108	12	P87760 hepatitis c
4	132	88.6	108	12	P87761 hepatitis c
5	132	88.6	415	12	Q81550 hepatitis c
6	131	87.9	73	12	Q8JVR2 hepatitis c
7	131	87.9	100	12	Q8QP82 hepatitis c
8	131	87.9	191	12	Q9IK25 hepatitis c
9	131	87.9	191	12	Q9IK27 hepatitis c
10	131	87.9	191	12	Q9IK26 hepatitis c
11	131	87.9	191	12	Q68413 hepatitis c
12	131	87.9	415	12	Q81266 hepatitis c
13	130	87.2	43	12	Q68305 hepatitis c
14	130	87.2	73	12	Q8JVR3 hepatitis c
15	130	87.2	109	12	Q81230 hepatitis c
16	130	87.2	109	12	Q81235 hepatitis c

17 130 87.2 109 12 Q81233 hepatitis c
18 130 87.2 109 12 Q81231 hepatitis c
19 130 87.2 109 12 Q81232 hepatitis c
20 130 87.2 109 12 Q81234 hepatitis c
21 130 87.2 109 12 Q81236 hepatitis c
22 130 87.2 150 12 Q68867 hepatitis c
23 130 87.2 150 12 Q68861 hepatitis c
24 130 87.2 150 12 Q68863 hepatitis c
25 130 87.2 150 12 Q68865 hepatitis c
26 130 87.2 416 12 Q81265 hepatitis c
27 130 87.2 782 12 Q68951 hepatitis c
28 129 86.6 109 12 Q81237 hepatitis c
29 127 85.2 109 12 Q81807 hepatitis c
30 126 84.6 45 12 Q68307 hepatitis c
31 126 84.6 45 12 Q68308 hepatitis c
32 126 84.6 45 12 Q68310 hepatitis c
33 126 84.6 60 12 Q8JVS2 hepatitis c
34 126 84.6 61 12 Q8JVS3 hepatitis c
35 126 84.6 73 12 Q8JVR4 hepatitis c
36 126 84.6 100 12 Q8QP85 hepatitis c
37 126 84.6 100 12 Q8QP87 hepatitis c
38 126 84.6 100 12 Q8QP72 hepatitis c
39 126 84.6 100 12 Q8QP71 hepatitis c
40 126 84.6 100 12 Q8QP74 hepatitis c
41 126 84.6 100 12 Q8QP80 hepatitis c
42 126 84.6 100 12 Q8QP77 hepatitis c
43 126 84.6 100 12 Q8QP84 hepatitis c
44 126 84.6 100 12 Q8QP75 hepatitis c
45 126 84.6 100 12 Q8QP79 hepatitis c

ALIGNMENTS

RESULT 1
Q81487 PRELIMINARY; PRT: 3023 AA.
ID Q81487;
AC Q81487;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus type 3b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=42791;
RN [1]
RP SEQUENCE FROM N.A.
RA Chayama K ;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D49374; BRA08372.1; -.
DR HSP; P27958; IALV.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR003521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.

DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS05057; RDRP_VIRAL; 1.
 DR PROSITE: PS05051; RDRP_VIRAL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 384
 FT CHAIN 385 737
 FT CHAIN 738 1014
 FT CHAIN 1015 1665
 FT CHAIN 1666 1980
 FT CHAIN 1981 3023
 SEQUENCE 3023 AA; 329734 MW; 5268D1EC410AC545 CRC64;
 Query Match 100.0%; Score 149; DB 12; Length 3023;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PKPKROTCKNTLRPRKNVFPAGGQIVG 28
 DB 5 PKPKROTCKNTLRPRKNVFPAGGQIVG 32
 RESULT 2
 ID Q81345 PRELIMINARY; PRT; 419 AA.
 AC Q81345;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN SEQUENCE FROM N.A.
 RP MEDLINE-94270990; PubMed-7545932;
 RA Hotta H., Doi H., Hayashi T., Purwanta M., Soemarto W., Mizokami M., Ohba K., Homma M.;
 "Analysis of the core and E1 envelope region sequences of a novel variant of hepatitis C virus obtained in Indonesia."; Arch. Virol. 136:53-62(1994).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-95088611; PubMed-7996156;
 RA Chayama K., Tsubota A., Koida I., Arase Y., Saitoh S., Ikeda K., Kumada H.;
 "Nucleotide sequence hepatitis C virus (type 3b) isolated from a Japanese patient with chronic hepatitis C."; J. Gen. Virol. 75:3623-3628(1994).
 RL CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D11443; BAA02007.1; -;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR000169; SHprot.acsite.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.

DR Pfam: PF01560; HCV_NSI; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; transmembrane.
 FT CHAIN 1 >190
 FT CHAIN 191 >419
 FT NON_TER 419 419
 SEQUENCE 419 AA; 45391 MW; CB8956E32EA6DD6C CRC64;
 Query Match 91.9%; Score 137; DB 12; Length 419;
 Best Local Similarity 89.3%; Pred. No. 1.6e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPKROTCKNTLRPRKNVFPAGGQIVG 28
 DB 5 PKPKROTCKNTLRPRKNVFPAGGQIVG 32
 RESULT 3
 ID P87760 PRELIMINARY; PRT; 108 AA.
 AC P87760;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Core protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN SEQUENCE FROM N.A.
 RA Ohno T., Mizokami M.;
 "Determination of nine genotypes of hepatitis C virus using PCR method."; Thesis (1995), Nagoya City University Medical School.
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-96305314; PubMed-8712927;
 RA Ohno T., Mizokami M., Saleh M.G., Orito E., Ohba K.-I., Wu R.-R., Koide T., Tibbs C.J., Nouri-Aria K.T., Tokudome S., Williams R.;
 "Usefulness and limitation of phylogenetic analysis for hepatitis C virus core region: application to isolates from Egyptian and Yemeni patients."; Arch. Virol. 141:1101-1113(1996).
 RL CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL: D82032; BAA11517.1; -;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 108 108
 SEQUENCE 108 AA; 11773 MW; A05A3836EC14BA21 CRC64;
 Query Match 88.6%; Score 132; DB 12; Length 108;
 Best Local Similarity 85.7%; Pred. No. 2.1e-12;
 Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPKROTCKNTLRPRKNVFPAGGQIVG 28
 DB 5 PKPKROTCKNTLRPRKNVFPAGGQIVG 32
 RESULT 4
 ID P87761 PRELIMINARY; PRT; 108 AA.
 AC P87761;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Core protein (Genome polyprotein) (Fragment).

```

OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohno T., Mizokami M.;
RT "determination of nine genotypes of hepatitis C virus using PCR
method.";
RL Thesis (1995), Nagoya City University Medical School.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305314; PubMed=97112927;
RA Ohno T., Mizokami M., Saleh M.G., Orito E., Ohba K.-I., Wu R.-R.,
RA Koide T., Tibbs C.J., Nouri-Aria K.T., Tokudome S., Williams R.;
RT "Usefulness and limitation of phylogenetic analysis for hepatitis C
RT virus core region: application to isolates from Egyptian and Yemeni
RT patients.";
RL Arch. Virol. 141:1101-1113(1996).
~C -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
~C LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
~C PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
~C PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D82033; BAA11518.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 11773 MW; A05A3836EC14BA21 CRC64;

Query Match 88.6%; Score 132; DB 12; Length 108;
Best Local Similarity 85.7%; Pred. No. 2.1e-12;
Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPKRQTKRNTLRPRKNVKEPAGGQIVG 28
DB 5 PKPQRTKRTTIRRPQNVKPPGGQIVG 32
|||||
|||||

RESULT 5
OY Q81550 PRELIMINARY; PRT; 415 AA.
AC Q81550;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-2002 (TrEMBLrel. 22, Last annotation update)
DE Core, env, and part of E2/NS1 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NE137;
RC MEDLINE=94186155; PubMed=8138250;
RA Shrestha S.M., Tsuda F., Okamoto H., Tokita H., Horikita M.,
RA Tanaka T., Miyakawa Y., Mayumi M.;
RT "Hepatitis B virus subtypes and hepatitis C virus genotypes in
RT patients with chronic liver disease in Nepal.";
RL Hepatology 19:805-809(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=NE137;
RC MEDLINE=94201770; PubMed=8151307;
RA Tokita H., Shrestha S.M., Okamoto H., Sakamoto M., Horikita M.,
RA Iizuka H., Shrestha S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Nepal with novel genotypes and their
RT classification into the third major group.";
RL J. Gen. Virol. 75:931-936(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).

```

```

DR EMBL; D16616; BAA04038.1; -
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01559; HCV_env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 415 415
SQ SEQUENCE 415 AA; 44703 MW; 0AD63410FA8F522F CRC64;

Query Match 88.6%; Score 132; DB 12; Length 415;
Best Local Similarity 85.7%; Pred. No. 9.1e-12;
Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPKRQTKRNTLRPRKNVKEPAGGQIVG 28
DB 5 PKPQRTKRTTIRRPQNVKPPGGQIVG 32
|||||
|||||

RESULT 6
OY Q8JYR2 PRELIMINARY; PRT; 73 AA.
AC Q8JYR2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RIG304;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF506623; AAM33399.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
FT NON_TER 1 1
FT NON_TER 73 73
SQ SEQUENCE 73 AA; 8272 MW; 64D60827A48FD796 CRC64;

Query Match 87.9%; Score 131; DB 12; Length 73;
Best Local Similarity 82.1%; Pred. No. 1.9e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPKRQTKRNTLRPRKNVKEPAGGQIVG 28
DB 1 PKPQRTKRTTIRRPQNVKPPGGQIVG 28
|||||
|||||

RESULT 7
OY Q8QP82 PRELIMINARY; PRT; 100 AA.
AC Q8QP82;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

```

```

OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=775;
RA MEDLINE=21904745; PubMed=11907242;
RX Kalina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
RL in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AY070183; AAL58597.1; -.
DR InterPro: IPR002522; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100 100
SQ SEQUENCE 100 AA; 11235 MW; F2A0FD95F8E85AEE CRC64;

Query Match 87.9%; Score 131; DB 12; Length 100;
Best Local Similarity 82.1%; Pred. No. 2.7e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKNTLRPRKNVFPAGGQIVG 28
DB 5 PKPQRTKTRTIRRPQNVKFPGGQIVG 32
||||:|||||:|||||:|||||

RESULT 8
Q9IK25 PRELIMINARY; PRT; 191 AA.
ID Q9IK25
AC Q9IK25;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core structural protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SOM3;
RX MEDLINE=20273994; PubMed=10811932;
RA Abid K., Quadri R., Veuthey A.L., Hadengue A., Negro F.;
RT "A novel hepatitis C virus (HCV) subtype from Somalia and its
RT classification in HCV clade 3.";
RL J. Gen. Virol. 81:1485-1493(2000).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF216794; AAF44741.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20730 MW; 7E85DDDE0DC4AC3A CRC64;

Query Match 87.9%; Score 131; DB 12; Length 191;
Best Local Similarity 82.1%; Pred. No. 5.5e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKNTLRPRKNVFPAGGQIVG 28
DB 5 PKPQRTKTRTIRRPQNVKFPGGQIVG 32
||||:|||||:|||||:|||||

RESULT 9
Q9IK27 PRELIMINARY; PRT; 191 AA.
ID Q9IK27
AC Q9IK27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core structural protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SOM1;
RX MEDLINE=20273994; PubMed=10811932;
RA Abid K., Quadri R., Veuthey A.L., Hadengue A., Negro F.;
RT "A novel hepatitis C virus (HCV) subtype from Somalia and its
RT classification in HCV clade 3.";
RL J. Gen. Virol. 81:1485-1493(2000).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF216792; AAF44739.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20730 MW; 7E85DDDE0DC4AC3A CRC64;

Query Match 87.9%; Score 131; DB 12; Length 191;
Best Local Similarity 82.1%; Pred. No. 5.5e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKNTLRPRKNVFPAGGQIVG 28
DB 5 PKPQRTKTRTIRRPQNVKFPGGQIVG 32
||||:|||||:|||||:|||||

RESULT 10
Q9IK26 PRELIMINARY; PRT; 191 AA.
ID Q9IK26
AC Q9IK26;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core structural protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SOM2;
RX MEDLINE=20273994; PubMed=10811932;
RA Abid K., Quadri R., Veuthey A.L., Hadengue A., Negro F.;
RT "A novel hepatitis C virus (HCV) subtype from Somalia and its
RT classification in HCV clade 3.";
RL J. Gen. Virol. 81:1485-1493(2000).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF216793; AAF44740.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20730 MW; FC844F8E1573F615 CRC64;
```

```

Query Match      87.9%; Score 131; DB 12; Length 191;
Best Local Similarity 82.1%; Pred. No. 5.5e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPPKNVKKFPAGGQIVG 28
    |||:|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTIRPPQNVKFPGGGQIVG 32

RESULT 11
Q68413 PRELIMINARY; PRT; 191 AA.
ID Q68413
AC Q68413
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=OC29;
RX MEDLINE=97052554; PubMed=8897188;
RA Bernier L., Willem B., Delage G., Murphy D.G.;
RT "Identification of numerous hepatitis C virus genotypes in Montreal,
Canada.";
RL J. Clin. Microbiol. 34:2815-2818(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; U33437; AAB40040.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191
FT SEQUENCE 191 AA; 20720 MW; 7A4B84532A1D07F9 CRC64;

Query Match      87.9%; Score 131; DB 12; Length 191;
Best Local Similarity 82.1%; Pred. No. 5.5e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPPKNVKKFPAGGQIVG 28
    |||:|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTIRPPQNVKFPGGGQIVG 32

RESULT 12
Q81266 PRELIMINARY; PRT; 415 AA.
ID Q81266
AC Q81266
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Th576;
RX MEDLINE=96005057; PubMed=7561773;
RA Tokita H., Okamoto H., Luengrojjanakul P., Vareesangthip K.,
Chaiyavati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Thailand classifiable into five novel
genotypes in the sixth (5b), seventh (7c,7d) and ninth (9b, 9c) major
genetic groups.";
RL J. Gen. Virol. 76:2329-2335(1995).

```

```

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D37840; BAA07087.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; transmembrane.
FT NON_TER 415
FT SEQUENCE 415 AA; 44881 MW; FD151EF15B1C31FC CRC64;

Query Match      87.9%; Score 131; DB 12; Length 415;
Best Local Similarity 85.7%; Pred. No. 1.3e-11;
Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPPKNVKKFPAGGQIVG 28
    |||:|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTIRPPQNVKFPGGGQIVG 32

RESULT 13
Q68305 PRELIMINARY; PRT; 43 AA.
ID Q68305
AC Q68305
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
NCBI_TaxID=11103;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-BB46;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DDJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; U23743; AAG65050.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 43
FT SEQUENCE 43 AA; 4844 MW; A173B6263AF6DC4F CRC64;

Query Match      87.2%; Score 130; DB 12; Length 43;
Best Local Similarity 82.1%; Pred. No. 1.5e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPPKNVKKFPAGGQIVG 28
    |||:|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTIRPPQNVKFPGGGQIVG 32

RESULT 14
Q8JYR3 PRELIMINARY; PRT; 73 AA.
ID Q8JYR3
AC Q8JYR3
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 23, Last annotation update)

```


DE Genome polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIG292;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RL "Genetic variability of hepatitis C virus in Western Siberia.";
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF506622; AAM33398.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
KW NON_TER 1 73
SEQUENCE 73 AA; 8258 MW; 76549827A490C896 CRC64;

Query Match 87.2%; Score 130; DB 12; Length 73;
Best Local Similarity 82.1%; Pred. No. 2.8e-12;
Matches 23; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKNTLRPRKNVFPAGGQIVG 28
|||:|||||:|||||:|||||
DB 1 PKPQTKNTVRPPONVFPGGQIVG 28

RESULT 15
Q81230 PRELIMINARY; PRT; 109 AA.
AC Q81230;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Core protein (Genome polypeptide) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Banl-1;
RA Ohno T., Mizokami M., Chstopher T., Ohba K., Suzuki K., Roger W.;
"The nucleotide sequence of the core region of HCV in Pakistan and
Bangladesh: The geographic characterization of HCV in South Asia.";
Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D29647; BAA21022.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
KW NON_TER 109 109
SEQUENCE 109 AA; 12293 MW; 524E8D425A7E3CB2 CRC64;

Query Match 87.2%; Score 130; DB 12; Length 109;
Best Local Similarity 85.7%; Pred. No. 4.3e-12;
Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKQTKNTLRPRKNVFPAGGQIVG 28
|||:|||||:|||||:|||||
DB 5 PKPQTKNTVRPPONVFPGGQIVG 32

Search completed: August 7, 2003, 11:19:03
Job time : 26.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-32

Perfect score: 149

Sequence: 1 PKPKROTGRNTLRPNKVPAGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

cal number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	100.0	28	3	US-08-921-887-32
2	137	91.9	319	4	US-08-635-886C-211
3	126	84.6	28	3	US-08-921-887-23
4	126	84.6	191	2	US-08-290-665A-187
5	126	84.6	191	2	US-08-290-665A-188
6	126	84.6	191	2	US-08-290-665A-189
7	126	84.6	191	5	PCT-US95-10398-187
8	126	84.6	191	5	PCT-US95-10398-188
9	126	84.6	191	5	PCT-US95-10398-190
10	126	84.6	319	4	US-08-836-075A-44
11	126	84.6	319	4	US-08-835-886C-230
12	125	83.9	28	3	US-08-921-887-24
13	125	83.9	74	3	US-08-836-075A-104
14	125	83.9	191	2	US-08-290-665A-189
15	125	83.9	191	5	PCT-US95-10398-189
16	123	82.6	28	3	US-08-921-887-25
17	123	82.6	450	4	US-08-635-886C-191
18	123	82.6	450	4	US-08-635-886C-192
19	122	81.9	191	2	US-08-290-665A-175
20	122	81.9	191	5	PCT-US95-10398-175
21	121	81.2	34	3	US-08-380-160-6
22	121	81.2	43	4	US-09-020-846-36
23	121	81.2	44	3	US-08-380-160-2
24	121	81.2	44	4	US-09-389-756-1
25	121	81.2	45	3	US-08-380-160-1
26	121	81.2	61	1	US-07-946-054-9
27	121	81.2	61	1	US-08-083-947-23

28 121 81.2 61 1 US-08-530-550-3
29 121 81.2 61 1 US-08-282-037-26
30 121 81.2 61 5 PCT-US93-08638-9
31 121 81.2 61 5 PCT-US94-07088-23
32 121 81.2 61 5 PCT-US95-13660-3
33 121 81.2 74 3 US-08-836-075A-10
34 121 81.2 74 4 US-08-635-886C-198
35 121 81.2 100 4 US-08-635-886C-232
36 121 81.2 108 3 US-08-836-075A-14
37 121 81.2 115 1 US-08-324-977-8
38 121 81.2 115 2 US-08-384-616-8
39 121 81.2 115 2 US-08-904-686A-8
40 121 81.2 115 3 US-09-315-850-8
41 121 81.2 123 2 US-08-501-195-2
42 121 81.2 137 3 US-08-836-075A-46
43 121 81.2 138 3 US-08-836-075A-60
44 121 81.2 140 2 US-08-501-195-1
45 121 81.2 154 3 US-08-854-531-2

ALIGNMENTS

RESULT 1
US-08-921-887-32
; Sequence 32, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; NUMBER OF INVENTIONS: 55
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-32

Query Match 100.0%; Score 149; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.3e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
US-08-921-887-23
; Sequence 23, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799

Query Match	84.6%	Score 126;	DB 2;	Length 191;
Best Local Similarity	78.6%	Pred. NO. 5.8e-11;		
Matches 22:	Conservative	5;	Mismatches 1;	Indels 0;
	Conservative	5;	Mismatches 1;	Indels 0;
	Conservative	5;	Mismatches 1;	Indels 0;

Qy 1 PKPKRQTKRNTLRRPKNVKFPAGGQIVG 28
 |||:||||:||||:||||:||||:||||:||||
 Db 5 PKPQRKTKRNTIRRPQDVKFPGGGQIVG 32

RESULT 5

```

US-08-290-665A-188
; Sequence 188, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R. H. AND
; APPLICANT: PURCELL, R. H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263

```

ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

```

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

```

; INFORMATION FOR SEQ ID NO: 188:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 191 amino acids

```

; amino acid
; STRANDEDNESS: unknown
: TOPOLOGY: unknown

ORIGINAL SOURCE:
ORGANISM: *homo sapiens*
INDIVIDUAL ISOLATE: S52

US-08-290-665A-188

Query Match	84.6%	Score 126;	DB 2;	Length 191;
Best Local Similarity	78.6%	Pred. No. 5.8e-11;		
Matches 22;	Conservative	5;	Mismatches 1;	Indels 0;
Gaps				

Qy 1 PKPKRQTKRNTLRRPKNVKFPAGGQIVG 28
 | | | : | | | | : | | | : | | | |
Dd 5 PKPQRKTKRNTIRRPQDVKFPGGGQIVG 32

RESULT 6

```

US-08-290-665A-190
:
: Sequence 190, Application US/08290665A
: Patent No. 582852
:
: GENERAL INFORMATION:
:
: APPLICANT: BUKH, J., MILLER, R.H. AND
: APPLICANT: PURCELL, R.H.
:
: TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
: TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
: TITLE OF INVENTION: CORE GENES OF HEPATITIS C VIRUS
:

```

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 NUMBER OF SEQUENCES: 263
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435

CLASSIFICATION: 933
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TEXT: 421792

```

, ID: 421752
, INFORMATION FOR SEQ ID NO: 190:
,-----
, SEQUENCE CHARACTERISTICS:
,     LENGTH: 191 amino acids
,     TYPE: amino acid
,     STRANDEDNESS: unknown
,     TOPOLOGY: unknown
,-----
, ORIGINAL SOURCE:
,     ORGANISM: homosapiens
,     INDIVIDUAL ISOLATE: DK12
,     US-08-290-665A-190

```

Query Match 84.6%; Score 126; DB 2; Length 191;
Best Local Similarity 78.6%; Pred. No. 5.8e-11;
Matches 22: Conservative 5; Mismatches 1; Indels

QY 1 PKPKRQTKRNTLRRPKNVKFPAGGQIVG 28
 IIII:IIII:III:III:IIII
 Db 5 PKPQKTKRNTIRRPQDVKFPGGGOIVG 32

RESULT 7

PCT-US95-10398-187
 ; Sequence 187, Application PC/TUS9510398
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R. H. AND
 ; APPLICANT: PURCELL, R. H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/10398

```

; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 187:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: HK10
PCT-US95-10398-187

```

```

Query Match      84.6%; Score 126; DB 5; Length 191;
Best Local Similarity 78.6%; Pred. No. 5.8e-11;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 PKPKRQTKRNTLRPRKNVFPAGGQIVG 28
    |||:|||||:|||||:|||||
Db 5 PKPQRTKRTIRRPQDVKFPGGQIVG 32

```

```

RESULT 8
PCT-US95-10398-188
; Sequence 188, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R. H. AND
; APPLICANT: PURCELL, R. H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459

```

```

; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S52
PCT-US95-10398-188

```

```

Query Match      84.6%; Score 126; DB 5; Length 191;
Best Local Similarity 78.6%; Pred. No. 5.8e-11;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 PKPKRQTKRNTLRPRKNVFPAGGQIVG 28
    |||:|||||:|||||:|||||
Db 5 PKPQRTKRTIRRPQDVKFPGGQIVG 32

```

```

RESULT 9
PCT-US95-10398-190
; Sequence 190, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R. H. AND
; APPLICANT: PURCELL, R. H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 190:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown

```

```
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK12
PCT-US95-10398-190

Query Match      84.6%; Score 126; DB 5; Length 191;
Best Local Similarity 78.6%; Pred. No. 5 8e-11;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPPKPKVFPAGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRKTRNTIRPDQVFPKPGGQIVG 32

RESULT 10
US-08-836-075A-44
; Sequence 44, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS: 207
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-836-075A-44

Query Match      84.6%; Score 126; DB 3; Length 319;
Best Local Similarity 78.6%; Pred. No. 1e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPPKPKVFPAGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRKTRNTIRPDQVFPKPGGQIVG 32

RESULT 11
US-08-635-886C-230
; Sequence 230, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 230
; LENGTH: 319
; TYPE: PPT
; ORGANISM: hepatitis C virus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (144)..(144)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (156)..(157)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (161)..(161)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (167)..(167)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (171)..(172)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (174)..(174)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (177)..(177)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (232)..(232)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (233)..(233)
; OTHER INFORMATION: Xaa is any amino acid
US-08-635-886C-230

Query Match      84.6%; Score 126; DB 4; Length 319;
Best Local Similarity 78.6%; Pred. No. 1e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPPKPKVFPAGGQIVG 28
   |||:|||||:|||||:|||||
Db 5 PKPQRKTRNTIRPDQVFPKPGGQIVG 32

RESULT 12
US-08-921-887-24
```

```
; Sequence 24, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-24

Query Match      83.9%; Score 125; DB 3; Length 28;
Best Local Similarity 52.1%; Pred. No. 1e-11;
Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||||| |||:||||| |||||
Db      1 PKPQRTKRNTRRPQDVKPPGGQIVG 28

RESULT 13
US-08-836-075A-104
; Sequence 104, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVEN, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P. O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:

; Sequence 24, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-24

Query Match      83.9%; Score 125; DB 3; Length 28;
Best Local Similarity 52.1%; Pred. No. 1e-11;
Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||||| |||:||||| |||||
Db      1 PKPQRTKRNTRRPQDVKPPGGQIVG 28

RESULT 14
US-08-290-665A-189
; Sequence 189, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA: US/08/836,075A
; APPLICATION NUMBER: 94870166.9
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-075A-104

Query Match      83.9%; Score 125; DB 3; Length 74;
Best Local Similarity 82.1%; Pred. No. 2.9e-11;
Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||||| |||:||||| |||||
Db      5 PKPQRTKRNTRRPQDVKPPGGQIVG 32

RESULT 14
US-08-290-665A-189
; Sequence 189, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
```

TELEX: 421792
INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-189

Query Match 83.9%; Score 125; DB 2; Length 191;
Best Local Similarity 75.0%; Pred. No. 8.1e-11;
Matches 21; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGOIVG 28
||||:||||:||||:||||:||||:||||
DB 5 PKPQRTKRTIRRPQDIKFPGGGOIVG 32

-SULT 15
PCT-US95-10398-189
; Sequence 189, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
PCT-US95-10398-189

Query Match 83.9%; Score 125; DB 5; Length 191;
Best Local Similarity 75.0%; Pred. No. 8.1e-11;
Matches 21; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGOIVG 28
||||:||||:||||:||||:||||:||||
DB 5 PKPQRTKRTIRRPQDIKFPGGGOIVG 32

Search completed: August 7, 2003, 11:23:53
Job time : 10.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146a-32

Perfect score: 149

Sequence: 1 PRPKRQTKRLRRPNKVPAGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

al number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap:*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap:*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pap:*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pap:*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pap:*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap:*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pap:*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap:*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap:*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	84.6	319	10	US-09-851-138-44
2	125	83.9	74	10	US-09-851-138-104
3	121	81.2	44	15	US-10-367-677-1
4	121	81.2	74	10	US-09-851-138-10
5	121	81.2	91	9	US-09-758-308-1
6	121	81.2	97	10	US-09-756-875-8
7	121	81.2	103	10	US-09-921-397-77
8	121	81.2	108	10	US-09-851-138-14
9	121	81.2	113	10	US-09-921-397-78
10	121	81.2	137	10	US-09-851-138-46
11	121	81.2	138	10	US-09-851-138-60
12	121	81.2	166	11	US-09-899-046-152
13	121	81.2	165	11	US-09-878-281-152
14	121	81.2	169	11	US-09-899-046-42
15	121	81.2	169	11	US-09-899-046-44

16	121	81.2	169	11	US-09-878-281-42	Sequence 42, Appl
17	121	81.2	189	11	US-09-878-281-44	Sequence 44, Appl
18	121	81.2	182	10	US-09-929-955-2	Sequence 2, Appl
19	121	81.2	182	14	US-10-104-966-2	Sequence 2, Appl
20	121	81.2	191	11	US-09-194-949-3	Sequence 3, Appl
21	121	81.2	318	10	US-09-851-138-76	Sequence 76, Appl
22	121	81.2	319	10	US-09-851-138-12	Sequence 12, Appl
23	121	81.2	319	10	US-09-851-138-18	Sequence 18, Appl
24	121	81.2	319	11	US-09-899-046-50	Sequence 50, Appl
25	121	81.2	319	11	US-09-899-046-52	Sequence 52, Appl
26	121	81.2	319	11	US-09-899-046-54	Sequence 54, Appl
27	121	81.2	319	11	US-09-899-046-144	Sequence 144, Appl
28	121	81.2	319	11	US-09-878-281-50	Sequence 50, Appl
29	121	81.2	319	11	US-09-878-281-52	Sequence 52, Appl
30	121	81.2	319	11	US-09-878-281-54	Sequence 54, Appl
31	121	81.2	319	11	US-09-878-281-144	Sequence 144, Appl
32	121	81.2	809	10	US-09-973-025-50	Sequence 50, Appl
33	121	81.2	809	11	US-09-899-303-50	Sequence 50, Appl
34	121	81.2	809	11	US-09-995-808-50	Sequence 50, Appl
35	121	81.2	809	11	US-09-995-860-50	Sequence 50, Appl
36	121	81.2	2894	10	US-09-941-611-23	Sequence 23, Appl
37	121	81.2	2894	15	US-10-044-995-23	Sequence 23, Appl
38	121	81.2	2985	15	US-10-259-275-40	Sequence 40, Appl
39	121	81.2	3011	9	US-09-742-659-4	Sequence 4, Appl
40	121	81.2	3011	10	US-09-952-572-9	Sequence 9, Appl
41	121	81.2	3011	10	US-09-929-955-1	Sequence 1, Appl
42	121	81.2	3011	10	US-09-747-419-20	Sequence 20, Appl
43	121	81.2	3011	11	US-09-891-894-3	Sequence 3, Appl
44	121	81.2	3011	14	US-10-104-966-1	Sequence 1, Appl
45	121	81.2	3011	15	US-10-259-275-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-09-851-138-44
; Sequence 44, Application US/09851138
; Publication No. US20020183508A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

STUVVER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO

AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P. O. BOX 4433

CITY: HOUSTON

STATE: TEXAS

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION NUMBER: US/09/851.138

APPLICATION NUMBER: 09-May-2001

FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>

APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994

APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 44:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-851-138-44

Query Match      84.6%; Score 126; DB 10; Length 319;
Best Local Similarity 78.6%; Pred. No. 5.5e-10;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
Db 5 PKPQKTKRNTLRPKNVKFPAGGQIVG 32

RESULT 2
US-09-851-138-104
; Sequence 104, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P. O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-09-851-138-104

Query Match      83.9%; Score 125; DB 10; Length 74;
Best Local Similarity 82.1%; Pred. No. 1.6e-10;
Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
Db 5 PKPQKTKRNTLRPKNVKFPAGGQIVG 32

RESULT 3
```

```

US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10/367,677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PPT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; TITLE: of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match      81.2%; Score 121; DB 15; Length 44;
Best Local Similarity 78.6%; Pred. No. 3.3e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
Db 4 PKPQKTKRNTLRPKNVKFPAGGQIVG 31

RESULT 4
US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOS
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P. O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-09-851-138-104

Query Match      83.9%; Score 125; DB 10; Length 74;
Best Local Similarity 82.1%; Pred. No. 1.6e-10;
Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKQTKRNTLRPKNVKFPAGGQIVG 28
Db 5 PKPQKTKRNTLRPKNVKFPAGGQIVG 32

RESULT 3
```

;
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-851-138-10

Query Match 81.2%; Score 121; DB 10; Length 74;
Best Local Similarity 78.6%; Pred. No. 5.8e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||| |||:||||| |||:
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 5
US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US20020090607A1
; GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 14114.034902
; CURRENT APPLICATION NUMBER: US/09/758.308
; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 81.2%; Score 121; DB 9; Length 91;
Best Local Similarity 78.6%; Pred. No. 7.3e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||| |||:||||| |||:
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 6
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:

;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756.875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 81.2%; Score 121; DB 10; Length 97;
Best Local Similarity 78.6%; Pred. No. 7.8e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||| |||:||||| |||:
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 7
US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-77

Query Match 81.2%; Score 121; DB 10; Length 103;
Best Local Similarity 78.6%; Pred. No. 8.3e-10;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPKRQTKRNTLRPKNVKPPAGGQIVG 28
|||:||||| |||:||||| |||:
Db 18 PKPQKTKRNTNRRPQDVKFPGGQIVG 45

RESULT 8
US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT

Query Match 81.2%; Score 121; DB 10; Length 113;
Best Local Similarity 78.6%; Pred. No. 9.2e-10;

Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28
||||:||||| |||:||||| |||||||
Db 5 PKPQRKTRNTNRPPQDVKFPGGQIVG 32

RESULT 10
US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO-
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46

Query Match 81.2%; Score 121; DB 10; Length 137;
Best Local Similarity 78.6%; Pred. No. 1.1e-09;
Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRPKNVKFPAGGQIVG 28
||||:||||| |||:||||| |||||||
Db 5 PKPQRKTRNTNRPPQDVKFPGGQIVG 32

RESULT 11
US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO-
; AGENTS
; NUMBER OF SEQUENCES: 207

Query Match 81.2%; Score 121; DB 11; Length 169;
 Best Local Similarity 78.6%; Pred. No. 1.4e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRRPKNVKFPAGGGQIVG 28
 |||:||||| |||:||||| |||
 Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

RESULT 15
 US-09-899-046-44
 : Sequence 44, Application US/09899046
 : Publication No. US20030008274A1
 : GENERAL INFORMATION:
 : APPLICANT:
 : TITLE OF INVENTION: New sequences of hepatitis C virus
 : TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
 : NUMBER OF SEQUENCES: 270
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/899,046
 : FILING DATE:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/362,455
 : FILING DATE:
 : INFORMATION FOR SEQ ID NO: 44:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 169 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-09-899-046-44

Query Match 81.2%; Score 121; DB 11; Length 169;
 Best Local Similarity 78.6%; Pred. No. 1.4e-09;
 Matches 22; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPKRQTKRNTLRRPKNVKFPAGGGQIVG 28
 |||:||||| |||:||||| |||
 Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32

Search completed: August 7, 2003, 12:01:14
 Job time : 14.3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds
(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146A-31

Perfect score: 150

Sequence: 1 PKPQKTERNTNRQDYRFSGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

al number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	100.0	28	20	AAV06681 NC mosaic protein
2	139	92.7	191	17	AAV06681 Hepatitis C virus
3	138	92.0	36	16	AAW06487 CN14 fragment of H
4	138	92.0	38	14	AAV06681 HCV capsid peptide
5	138	92.0	38	14	AAV06681 HCV capsid peptide
6	138	92.0	38	14	AAV06681 HCV capsid peptide
7	138	92.0	38	15	AAV06681 Non-A, non-B hepat
8	138	92.0	43	19	AAV06681 Hepatitis C virus
9	138	92.0	44	19	AAV06681 Hepatitis C virus

10	138	92.0	44	20	AAV26952 Hepatitis C virus
11	138	92.0	44	21	AAV26952 Human hepatitis C
12	138	92.0	45	21	AAV26952 Human hepatitis C
13	138	92.0	55	13	AAV26952 HCV core-envelope
14	138	92.0	55	13	AAV26952 HCV core-envelope
15	138	92.0	55	13	AAV26952 HCV core-envelope
16	138	92.0	55	13	AAV26952 HCV core-envelope
17	138	92.0	55	13	AAV26952 HCV core-envelope
18	138	92.0	57	13	AAV26952 Non-A, Non-B Hepat
19	138	92.0	61	13	AAV26952 Peptide VIIIE base
20	138	92.0	61	16	AAV26952 Anti-HCV antibody
21	138	92.0	61	17	AAV26952 Prototypic peptide
22	138	92.0	61	18	AAV26952 HCV core protein p
23	138	92.0	66	12	AAV26952 PT-NANB viral stru
24	138	92.0	74	17	AAV26952 Hepatitis C virus
25	138	92.0	78	13	AAV26952 Antigen pHC101
26	138	92.0	78	13	AAV26952 Antigen pHC101
27	138	92.0	79	14	AAV26952 HCV fragment 1 / I
28	138	92.0	79	14	AAV26952 HCV fragment 2 / I
29	138	92.0	80	15	AAV26952 Branched peptide H
30	138	92.0	90	16	AAV26952 Hepatitis C virus
31	138	92.0	91	22	AAV26952 Antigenic epitope
32	138	92.0	97	14	AAV26952 HCV core protein N
33	138	92.0	97	16	AAV26952 Hepatitis C virus
34	138	92.0	97	20	AAV26952 Protein encoded by
35	138	92.0	97	20	AAV26952 Hepatitis C virus
36	138	92.0	103	23	AAV26952 HCV bait polypepti
37	138	92.0	105	21	AAV26952 Protein encoded by
38	138	92.0	108	17	AAV26952 Hepatitis C virus
39	138	92.0	113	23	AAV26952 HCV bait polypepti
40	138	92.0	114	17	AAV26952 HCV capsid core pr
41	138	92.0	115	13	AAV26952 HCV core-envelope
42	138	92.0	115	17	AAV26952 DNA fragment HC360
43	138	92.0	123	13	AAV26952 C11-21 NANBH-speci
44	138	92.0	124	21	AAV26952 Hepatitis type C v
45	138	92.0	137	17	AAV26952 Hepatitis C virus

ALIGNMENTS

RESULT 1

AAV06681
ID AAV06681 standard; Protein: 28 AA.

AC AAV06681;

XX 17-JUN-1999 (first entry)

DT NC mosaic protein amino acid fragment I.

DE Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
restriction endonuclease assisted ligation; vaccination.

XX Hepatitis C virus.

PN WO9910506-A1.

XX 04-MAR-1999.

PF 21-AUG-1998; 98WO-US17385.

PR 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1999-204671/17.

XX New mosaic protein, comprising a plurality of homologous antigenic
peptides from different genotypes of a species - useful for
detecting hepatitis infection in an individual

XX PS Claim 5; Fig 9; 66pp; English.

XX CC The invention relates to a mosaic protein, comprising a plurality of

CC homologous antigenic peptides from different genotypes of a species. The

CC antigenic peptides are from nucleocapsid (NC) proteins. A method for

CC synthesizing an artificial gene that encodes the mosaic protein is also

CC provided. The method is designated restriction endonuclease assisted

CC ligation (REAL). The mosaic protein and the artificial mosaic protein

CC are useful for detecting a hepatitis infection in an individual. The

CC mosaic gene and protein is also useful for vaccination against

CC infection, especially hepatitis C. The method of synthesizing the

CC artificial gene and the resulting mosaic protein improve the sensitivity,

CC spectrum of immunoreactivity, and antigen specificity of enzyme

CC immunoassays. This provides improved detection of hepatitis C virus.

CC Sequences AA06673-683 represent amino acid sequence of each monomer

CC comprising the NC mosaic protein.

XX SQ Sequence 28 AA;

Query Match 100.0%; Score 150; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 2.2e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

DB 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

RESULT 2

AA092985

ID AAR92985 standard; Protein; 191 AA.

XX AC AAR92985;

XX DT 02-OCT-1996 (first entry)

XX DE Hepatitis C virus isolate SA6 core protein.

XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX KW hepatitis.

XX OS Hepatitis C virus.

XX PN WO9605315-A2.

XX PD 22-FEB-1996.

XX PR 15-AUG-1995; 95WO-US10398.

XX PS 15-AUG-1994; 94US-0290665.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX PR WPI; 1996-139709/14.

XX DR N-PSDB; AAT16659.

PT DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

XX Claim 4; Page 221-222; 340pp; English.

XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers

CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX SQ Sequence 191 AA;

Query Match 92.7%; Score 139; DB 17; Length 191;

Best Local Similarity 89.3%; Pred. No. 6.5e-12;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

DB 5 PKPQKTERNTNRRPQDVRFSGGGQIVG 32

RESULT 3

AAW06487

ID AAW06487 standard; Protein; 36 AA.

XX AC AAW06487;

XX DT 31-JAN-1997 (first entry)

XX DE CN14 fragment of HCV core region.

XX KW CN14; CP14; core region; hepatitis C virus; HCV; detection;

XX KW antibody.

XX OS Synthetic.

XX PN JP06327482-A.

XX PD 29-NOV-1994.

XX PF 21-MAY-1993; 93JP-0156026.

XX PR 21-MAY-1993; 93JP-0156026.

XX PA (IMMO) IMMUNO JAPAN KK.

XX DR WPI; 1995-047903/07.

XX DR N-PSDB; AAT45055.

PT Detection of hepatitis C virus - using oligopeptide fragment of HCV

PT core region

XX PS Claim 1; Page 6; 7pp; Japanese.

XX CC This sequence is encoded by the oligonucleotide, CN14, and represents

CC the peptide fragment CP14. CP14 is a fragment of the core region

CC of hepatitis C virus (HCV). CP14 may be used in the detection of

CC HCV infection and to raise antibodies against it.

XX SQ Sequence 36 AA;

Query Match 92.0%; Score 138; DB 16; Length 36;

Best Local Similarity 89.3%; Pred. No. 1.5e-12;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

DB 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

RESULT 4

AAR30687

ID AAR30687 standard; peptide; 38 AA.

XX AC AAR30687;

XX DT 25-MAR-2003 (updated)

XX PR 11-MAY-1993 (first entry)

XX DE HCV capsid peptide No. 23.

XX KW Hepatitis C virus; HCV; open reading frame; "common" sequence;

CC The sequences given in AAR30665-89 represent fragments of the
 CC Hepatitis C virus (HCV) amino acid sequence. They represent the
 CC beginning of the HCV open reading frame to amino acid 38 and
 CC encompass the "common" sequence. These peptides are contained in
 CC the capsid protein of the virus and themselves contain epitope
 CC groups. These peptides can be used in immunoassays for HCV
 CC antibodies. In the diagnosis of non-A, non-B hepatitis (NANBH), and
 CC in competitive inhibition assay for detecting HCV specific
 CC antibodies.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 38 AA;

Query Match 92.0%; Score 138; DB 14; Length 38;
 Best Local Similarity 89.3%; Pred. No. 1.6e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 DB |||||||:|||||:|||||:|||||:|||||
 5 PKPQKTRKNTNRRPDVRFSGGGQIVG 32

---SULT 7

AAR54065
 ID AAR54065 standard; Protein; 38 AA.

XX
 AC AAR54065;

XX
 DT 14-FEB-1995 (first entry)

XX
 DE Non-A, non-B hepatitis virus coreI region fragment.

XX
 KW Non-A, non-B hepatitis virus; NANBH; hepatitis C virus; HCV;
 KW core; ENV; NS1; NS2; NS3; antigen; detection.

OS Hepatitis C virus.

XX
 PN JP06141870-A.

XX
 PD 24-MAY-1994.

XX
 PF 12-MAR-1992; 92JP-0088140.

XX
 PR 12-MAR-1992; 92JP-0088140.

XX
 PA (SANW) SANWA KAGAKU KENKYUSHO CO.

XX
 PA (TOFU) TONEN CORP.

XX
 PA (TOXR-) ZH TOKYO TO RINSHO IGAKU SOGO KENKYUSHO.

XX
 WPI: 1994-205028/25.

XX
 N-PSDB; AAQ64067.

XX
 PT DNA coding a Non-A, non-B hepatitis virus antigen - useful for
 PT detecting HCV within serum

XX
 PS Disclosure; Page 11; 22pp; Japanese.

XX
 CC Hepatitis C virus #4 and #6 genes were isolated (AAQ64068-69).

XX
 CC Both genes contain the core, ENV, NS1, NS2 and NS3 regions.

XX
 CC A core region fragment is given in AAQ64067.

XX
 SQ Sequence 38 AA;

Query Match 92.0%; Score 138; DB 15; Length 38;
 Best Local Similarity 89.3%; Pred. No. 1.6e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 DB |||||||:|||||:|||||:|||||:|||||
 5 PKPQKTRKNTNRRPDVRFSGGGQIVG 32

RESULT 8

AAW37380
 ID AAW37380 standard; Protein; 43 AA.

XX
 AC AAW37380;

XX
 DT 11-MAR-1998 (first entry)

XX
 DE Hepatitis C virus C-1 protein 1-43.

XX
 KW Hepatitis C virus; HCV; chimeric; antigen; detection; core region;
 KW epitope; NS3; NS4; infection.

XX
 OS Hepatitis c virus.

XX
 PN JP09278794-A.

XX
 PD 28-OCT-1997.

XX
 PF 10-FEB-1997; 97JP-0027015.

XX
 PR 09-FEB-1996; 96JP-0024045.

XX
 PA (TOFU) TONEN CORP.

XX
 DR WPI: 1998-022248/03.

XX
 PT New chimeric peptide antigen derived from hepatitis C virus protein
 PT - useful for detecting HCV infections

XX
 PS Disclosure; Page 24; 30pp; Japanese.

XX
 CC The present sequence represents a Hepatitis C virus (HCV) protein
 CC sequence from the disclosure of the present specification. The
 CC present specification describes a chimeric HCV peptide antigen which
 CC comprises at least 2 peptide epitope regions from the HCV polypeptide
 CC core region, 2 peptide epitope regions from the NS3 region and at
 CC least 2 peptide epitope regions from the NS4 region. The antigen binds
 CC specifically with an antibody produced by a human infected by HCV. The
 CC peptide can detect a wide range of HCV infections with high sensitivity.

XX
 SQ Sequence 43 AA;

Query Match 92.0%; Score 138; DB 19; Length 43;
 Best Local Similarity 89.3%; Pred. No. 1.8e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 DB |||||||:|||||:|||||:|||||:|||||
 5 PKPQKTRKNTNRRPDVRFSGGGQIVG 32

RESULT 9

AAW66083
 ID AAW66083 standard; peptide; 44 AA.

XX
 AC AAW66083;

XX
 DT 16-NOV-1998 (first entry)

XX
 DE Hepatitis C virus p21 protein N-terminal fragment.

XX
 KW antigenic; nucleocapsid; p21 protein; alpha-helical; immunogen;
 KW antibody; hepatitis C virus; epitope; N-terminal fragment.

XX
 OS hepatitis c virus.

XX
 PN WO9839360-A1.

XX
 PD 11-SEP-1998.

XX
 PF 05-MAR-1998; 98WO-FR00442.

XX
 PR 05-MAR-1997; 97FR-0002878.

QY 1 PKPQKTKRNTNRPPQDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 4 PKPQKTKRNTNRPPQDVRFSGGGQIVG 31

RESULT 12

AA94409
 ID AAY94409 standard; peptide; 45 AA.

XX AC AAY94409;

XX DT 11-SEP-2000 (first entry)

XX DE Human hepatitis C virus core protein N-terminus, residues 1-45.

XX KW Human; hepatitis C virus; HCV; immunogen; antigen; antibody; virucide;
 KW hepatotropic; anti-inflammatory; virus detection; vaccine.

XX OS Hepatitis C virus.

XX WO200031130-A1.

-- 02-JUN-2000.

PF 19-NOV-1999; 99WO-1B01933.

PR 20-NOV-1998; 98US-0196155.

XX PA (INNR) BIO MERIEUX.

PI Dalbon P, Jollivet M, Jollivet-Reynaud C;

XX WPI: 2000-411934/35.

XX PT Polypeptides that bind to anti-hepatitis C virus antibodies, useful for
 PT diagnosing and preventing hepatitis C infections.

PS Claim 2; Page 42; 50pp; English.

XX The present sequence is the 45 N-terminal amino acids of the core
 CC protein of the human hepatitis C virus (HCV). The sequence is an
 CC immunodominant region containing conformational type epitopes and linear
 CC type epitopes. It manifests an immunoreactivity with all the sera of
 CC individuals or blood samples infected with HCV and which possess
 CC antibodies directed against the core protein. An amino acid of the
 CC present sequence may be substituted for homologous amino acids and side
 CC chains and peptide bonds may be modified. For example, L-amino acids may
 CC be replaced by D-amino acids, amine groups may be acetylated, and so on.
 CC The present antigenic sequence and its antigenic derivatives may be used
 CC for detection of hepatitis C virus and for raising antibodies against
 CC the virus.

XA Sequence 45 AA;

Query Match 92.0%; Score 138; DB 21; Length 45;

Best Local Similarity 89.3%; Pred. No. 1.9e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTNRPPQDVRFSGGGQIVG 28

Db 5 PKPQKTKRNTNRPPQDVRFSGGGQIVG 32

RESULT 13

AA29534
 ID AAR29534 standard; protein; 55 AA.

XX AC AAR29534;

XX DT 25-MAR-2003 (updated)

DT 26-APR-1993 (first entry)

XX HCV core-envelope peptide N1-1.

XX

KW Clone; polypeptide; core-envelope; region; Hepatitis C; Virus; HCV;
 KW HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;
 KW suppress; control; proteolytic processing; precursor.

XX Hepatitis C virus.

XX EP518313-A2.

XX 16-DEC-1992.

XX 11-JUN-1992; 92EP-0109812.

XX 11-JUN-1991; 91JP-0139268.

PR 12-JUL-1991; 91JP-0172794.

PR 07-OCT-1991; 91JP-0287008.

PR 16-DEC-1991; 91JP-0332329.

PR 20-APR-1992; 92JP-0099957.

XX (MITU) MITSUBISHI KASEI CORP.

XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

PI Teranishi Y;

XX WPI: 1992-417213/51.

DR N-PSDB; AAQ32444.

XX New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections

PS Claim 5; Page 74; 305pp; English.

XX The sequences given in AAR29529-30 and AAR29534-42 represent various
 CC clones of the polypeptide core-envelope region of the Hepatitis C
 CC Virus (HCV) gene of the invention. These sequences were isolated from
 CC the serum of a patient suffering from hepatitis C (HC). The RNA
 CC sequences encoding these peptides were converted into cDNA using
 CC transcriptase in the presence of one of the primer sequences given
 CC in AAR29540-46. The cDNA sequences isolated represent different
 CC alleles of the same region of the HCV gene. The entire HCV protein
 CC (see AAR29527) is useful in the development of a diagnostic method
 CC which is more accurate and effective than conventional ones, in the
 CC detection of antibodies raised against a wide range of HCVs which
 CC have been hardly detected before. The complete gene may be used in
 CC an in vitro screening system for a substance capable of specifically
 CC suppressing or controlling a proteolytic processing of a precursor
 CC polypeptide of HCV.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 55 AA;

Query Match 92.0%; Score 138; DB 13; Length 55;

Best Local Similarity 89.3%; Pred. No. 2.3e-12;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTNRPPQDVRFSGGGQIVG 28

Db 5 PKPQKTKRNTNRPPQDVRFSGGGQIVG 32

RESULT 14

AA29535
 ID AAR29535 standard; protein; 55 AA.

XX AC AAR29535;

XX DT 25-MAR-2003 (updated)

DT 26-APR-1993 (first entry)

XX HCV core-envelope peptide N1-2.

KW Clone; polypeptide; core-envelope; region; Hepatitis C; Virus; HCV;
 KW HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;

KW suppress; control; proteolytic processing; precursor.

XX Hepatitis C virus.

PN EP518313-A2.

PD 16-DEC-1992.

XX 11-JUN-1992; 92EP-0109812.

XX 11-JUN-1991; 91JP-0139268.

PR 12-JUL-1991; 91JP-0172794.

PR 07-OCT-1991; 91JP-0287008.

PR 16-DEC-1991; 91JP-0332329.

PR 20-APR-1992; 92JP-0099957.

XX (MITU) MITSUBISHI KASEI CORP.

XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

PI Teranishi Y;

WPI; 1992-417213/51.

N-PSDB; AAQ32445.

XX New hepatitis C virus gene and its encoded protein - used for

diagnosing and vaccinating against hepatitis C virus infections

XX Claim 33; Page 87; 305pp; English.

XX The sequences given in AAR25929-30 and AAR29534-42 represent various

clones of the polypeptide core-envelope region of the Hepatitis C

Virus (HCV) gene of the invention. These sequences were isolated from

the serum of a patient suffering from hepatitis C (HC). The RNA

sequences encoding these peptides were converted into cDNA using

transcriptase in the presence of one of the primer sequences given

in AAQ32540-46. The cDNA sequences isolated represent different

alleles of the same region of the HCV gene. The entire HCV protein

(see AAR29527) is useful in the development of a diagnostic method

which is more accurate and effective than conventional ones, in the

detection of antibodies raised against a wide range of HCVs which

have been hardly detected before. The complete gene may be used in

an in vitro screening system for a substance capable of specifically

suppressing or controlling a proteolytic processing of a precursor

polypeptide of HCV.

XX (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 55 AA;

Query Match 92.0%; Score 138; DB 13; Length 55;

Best Local Similarity 89.3%; Pred. No. 2.3e-12;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRPPQDVRFSGGQIVG 28

|||||:|||||:|||||:|||||

Db 5 PKPQKTRNTNRPPQDVRFSGGQIVG 32

|||||:|||||:|||||:|||||

RESULT 15

AAR29536

ID AAR29536 standard; Protein; 55 AA.

XX AAR29536;

XX 25-MAR-2003 (updated)

DT 26-APR-1993 (first entry)

XX HCV core-envelope peptide S1-3.

XX Clone; polypeptide; core-envelope; region; Hepatitis C; Virus; HCV;

KW HC; cDNA; transcriptase; primer; allele; diagnostic; antibody;

KW suppress; control; proteolytic processing; precursor.

XX Hepatitis C virus.

XX

PN EP518313-A2.

XX 16-DEC-1992.

XX 11-JUN-1992; 92EP-0109812.

XX 11-JUN-1991; 91JP-0139268.

PR 12-JUL-1991; 91JP-0172794.

PR 07-OCT-1991; 91JP-0287008.

PR 16-DEC-1991; 91JP-0332329.

PR 20-APR-1992; 92JP-0099957.

XX (MITU) MITSUBISHI KASEI CORP.

XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

PI Teranishi Y;

WPI; 1992-417213/51.

N-PSDB; AAQ32446.

XX New hepatitis C virus gene and its encoded protein - used for

diagnosing and vaccinating against hepatitis C virus infections

XX Claim 33; Page 88-89; 305pp; English.

XX The sequences given in AAR25929-30 and AAR29534-42 represent various

clones of the polypeptide core-envelope region of the Hepatitis C

Virus (HCV) gene of the invention. These sequences were isolated from

the serum of a patient suffering from hepatitis C (HC). The RNA

sequences encoding these peptides were converted into cDNA using

transcriptase in the presence of one of the primer sequences given

in AAQ32540-46. The cDNA sequences isolated represent different

alleles of the same region of the HCV gene. The entire HCV protein

(see AAR29527) is useful in the development of a diagnostic method

which is more accurate and effective than conventional ones, in the

detection of antibodies raised against a wide range of HCVs which

have been hardly detected before. The complete gene may be used in

an in vitro screening system for a substance capable of specifically

suppressing or controlling a proteolytic processing of a precursor

polypeptide of HCV.

XX (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 55 AA;

Query Match 92.0%; Score 138; DB 13; Length 55;

Best Local Similarity 89.3%; Pred. No. 2.3e-12;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRPPQDVRFSGGQIVG 28

|||||:|||||:|||||:|||||

Db 5 PKPQKTRNTNRPPQDVRFSGGQIVG 32

|||||:|||||:|||||:|||||

Search completed: August 7, 2003, 11:14:09

Job time : 38.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-31

Perfect score: 150

Sequence: 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

al number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.76:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	ID	Description
1	138	92.0	108	2 S41353	genome polyprotein
2	138	92.0	108	2 S41355	genome polyprotein
3	138	92.0	108	2 S41357	genome polyprotein
4	138	92.0	108	2 S41348	genome polyprotein
5	138	92.0	112	2 S41371	genome polyprotein
6	138	92.0	112	2 S41341	genome polyprotein
7	138	92.0	114	2 S41370	genome polyprotein
8	138	92.0	114	2 S41369	genome polyprotein
9	138	92.0	114	2 S41368	genome polyprotein
10	138	92.0	115	2 S41342	genome polyprotein
11	138	92.0	115	2 S41344	genome polyprotein
12	138	92.0	115	2 S41350	genome polyprotein
13	138	92.0	115	2 S41354	genome polyprotein
14	138	92.0	115	2 S41345	genome polyprotein
15	138	92.0	115	2 S41347	genome polyprotein
16	138	92.0	115	2 S41343	genome polyprotein
17	138	92.0	118	2 S41346	genome polyprotein
18	138	92.0	359	2 S21471	genome polyprotein
19	138	92.0	441	2 S12707	genome polyprotein
20	138	92.0	513	2 PC1284	genome polyprotein
21	138	92.0	520	2 JQ1925	polyprotein - hepa
22	138	92.0	523	2 JQ1926	genome polyprotein
23	138	92.0	550	2 JH0711	genome polyprotein
24	138	92.0	782	2 S19876	genome polyprotein
25	138	92.0	782	2 S18031	genome polyprotein
26	138	92.0	782	2 S18032	genome polyprotein
27	138	92.0	787	2 PN0677	hypothetical prote
28	138	92.0	874	2 JQ0883	genome polyprotein
29	138	92.0	874	2 JQ0881	genome polyprotein

30 138 92.0 876 2 PC2219 polypeptide - hepa
31 138 92.0 3010 1 GNWVTC genome polyprotein
32 138 92.0 3010 1 GNWVTC genome polyprotein
33 138 92.0 3010 1 S18030 genome polyprotein
34 138 92.0 3011 1 GNWVCH genome polyprotein
35 138 92.0 3011 1 S40770 genome polyprotein
36 138 92.0 3014 1 JC5620 genome polyprotein
37 138 92.0 3033 1 JQ1303 genome polyprotein
38 138 92.0 3033 1 GNWVJ8 genome polyprotein
39 137 91.3 125 2 S41352 genome polyprotein
40 134 89.3 115 2 S41351 genome polyprotein
41 134 89.3 115 2 S41349 genome polyprotein
42 134 89.3 640 2 JQ1584 genome polyprotein
43 133 88.7 88 2 S21336 genome polyprotein
44 133 88.7 109 2 S41367 genome polyprotein
45 133 88.7 114 2 S41365 genome polyprotein

ALIGNMENTS

RESULT 1

S41353

genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 2, N2

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41353

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.

A:Reference number: S41341

A:Accession: S41353

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229456

A:Experimental source: genotype 2, N2

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 108;

Best Local Similarity 89.3%; Pred. No. 2.1e-12;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28

|||||:|||||:|||||:|||||

Db 5 PKPQKTKNTNRRPQDVRFSGGGQIVG 32

RESULT 2

S41355

genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 2, N4

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41355

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.

A:Reference number: S41341

A:Accession: S41355

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229458

A:Experimental source: genotype 2, N4

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 108;

Best Local Similarity 89.3%; Pred. No. 2.1e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRRPQDVRFSGGGQIVG 32

RESULT 3

S41357 genome polypeptide - hepatitis C virus (genotype 2, N6) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41357
R:van Doorn, L.J.; Kletter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

ccession: S41357

olecule type: genomic RNA

...residues: 1-108 <VAN>

A:Cross-references: EMBL:229460

A:Experimental source: genotype 2, N6

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; polypeptide

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 108;

Best Local Similarity 89.3%; Pred. No. 2.1e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRRPQDVRFSGGGQIVG 32

RESULT 4

S41348 genome polypeptide - hepatitis C virus (genotype 1, N6) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41348
R:van Doorn, L.J.; Kletter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

ccession: S41348

olecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229451

A:Experimental source: genotype 1, N6

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; polypeptide

F:1-108/Product: core protein #status predicted <MAT>

Query Match

Best Local Similarity 92.0%; Score 138; DB 2; Length 108;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRRPQDVRFSGGGQIVG 32

RESULT 5

S41371

genome polypeptide - hepatitis C virus (genotype 5, N5) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus

A:Variety: genotype 5, N5
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41371
R:van Doorn, L.J.; Kletter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

A:Accession: S41371

A:Molecule type: genomic RNA

A:Residues: 1-112 <VAN>

A:Cross-references: EMBL:229474

A:Experimental source: genotype 5, N5

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; polypeptide

F:1-112/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 112;

Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRRPQDVRFSGGGQIVG 32

RESULT 6

S41341

genome polypeptide - hepatitis C virus (genotype 1, N1) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 1, N1

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41341

R:van Doorn, L.J.; Kletter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41341

A:Molecule type: genomic RNA

A:Residues: 1-112 <VAN>

A:Cross-references: EMBL:229444; NID:q443850; PIDN:CAA82582.1; PID:q443851

A:Experimental source: genotype 1, N1

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; polypeptide

F:1-112/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 112;

Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRRPQDVRFSGGGQIVG 32

RESULT 7

S41370

genome polypeptide - hepatitis C virus (genotype 5, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 5, N4

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41370

R:van Doorn, L.J.; Kletter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41370

A:Molecule type: genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: EMBL:229473; NID:q443908; PIDN:CAA82611.1; PID:q443909

A:Experimental source: genotype 5, N4

C:Superfamily: hepatitis C virus genome polypeptide

C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 8

S41369
genome polyprotein - hepatitis C virus (genotype 5, N3) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 5, N3
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
Accession: S41369

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41369
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>
A;Cross-references: EMBL:Z29472; NID:g443906; PIDN:CAA82610.1; PID:g443907
A;Experimental source: genotype 5, N3
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 9

S41368
genome polyprotein - hepatitis C virus (genotype 5, N2) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 5, N2
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
Accession: S41368

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41368
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>
A;Cross-references: EMBL:Z29471; NID:g443904; PIDN:CAA82609.1; PID:g443905
A;Experimental source: genotype 5, N2
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 10

S41342

genome polyprotein - hepatitis C virus (genotype 1, N10) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 1, N10
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S41342
R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41342
A;Molecule type: genomic RNA
A;Residues: 1-115 <VAN>
A;Cross-references: EMBL:Z29445; NID:g443852; PIDN:CAA82583.1; PID:g443853
A;Experimental source: genotype 1, N10
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-115/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 11

S41344
genome polyprotein - hepatitis C virus (genotype 1, N2) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 1, N2
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S41344

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41344
A;Molecule type: genomic RNA
A;Residues: 1-115 <VAN>
A;Cross-references: EMBL:Z29447; NID:g443856; PIDN:CAA82585.1; PID:g443857
A;Experimental source: genotype 1, N2
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-115/Product: core protein #status predicted <MAT>

Query Match 92.0%; Score 138; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 12

S41350
genome polyprotein - hepatitis C virus (genotype 1, N8) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 1, N8
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S41350

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41350
A;Molecule type: genomic RNA


```

A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229453; NID:g443868; PIDN:CAA82591.1; PID:g443869
A:Experimental source: genotype 1, N8
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F.1-115/Product: core protein #status predicted <MAT>

Query Match          92.0%; Score 138; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPDVRFSGGGQIVG 28
    |||||:|||||:|||||:| |||||
Db 5 PKPQKTRNTNRRPDVRFSGGGQIVG 32

RESULT 13
S41354
genome polyprotein - hepatitis C virus (genotype 2, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41354
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41354
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229457
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F.1-115/Product: core protein #status predicted <MAT>

Query Match          92.0%; Score 138; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPDVRFSGGGQIVG 28
    |||||:|||||:|||||:| |||||
Db 5 PKPQKTRNTNRRPDVRFSGGGQIVG 32

RESULT 14
S41345
genome polyprotein - hepatitis C virus (genotype 1, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41345
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41345
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229448; NID:g443858; PIDN:CAA82586.1; PID:g443859
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F.1-115/Product: core protein #status predicted <MAT>

Query Match          92.0%; Score 138; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPDVRFSGGGQIVG 28
    |||||:|||||:|||||:| |||||

```

```

Db 5 PKPQKTRNTNRRPDVRFSGGGQIVG 32

RESULT 15
S41347
genome polyprotein - hepatitis C virus (genotype 1, N5) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N5
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41347
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41347
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229450; NID:g443862; PIDN:CAA82588.1; PID:g443863
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F.1-115/Product: core protein #status predicted <MAT>

Query Match          92.0%; Score 138; DB 2; Length 115;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPDVRFSGGGQIVG 28
    |||||:|||||:|||||:| |||||
Db 5 PKPQKTRNTNRRPDVRFSGGGQIVG 32

Search completed: August 7, 2003, 11:21:49
Job time: 10.5455 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146a-31

Perfect score: 150

Sequence: 1 PKPQKTEENTNRQDVRFSGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

al number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138	92.0	513	1	POLG_HCVJ2
2	138	92.0	520	1	POLG_HCVH4
3	138	92.0	520	1	POLG_HCVHK
4	138	92.0	737	1	POLG_HCVJ5
5	138	92.0	737	1	POLG_HCVJ7
6	138	92.0	3010	1	POLG_HCVBK
7	138	92.0	3010	1	POLG_HCVJA
8	138	92.0	3011	1	POLG_HCVH
9	138	92.0	3033	1	POLG_HCVJ6
10	138	92.0	3033	1	POLG_HCVJ8
11	131	87.3	3010	1	POLG_HCVTW
12	130	86.7	3011	1	POLG_HCVJT
13	130	86.7	3011	1	POLG_HCV1
14	53.5	35.7	602	1	2A5D_HUMAN
15	52.5	35.0	586	1	2A5D_RABIT
16	52	34.7	512	1	IE63_HSV2H
17	50	33.3	622	1	ESR1_ONCMY
18	49.5	33.0	303	1	YB1_XENLA
19	48	32.0	282	1	RK4_TOBAC
20	48	32.0	349	1	SCD6_YEAST
21	47	31.3	551	1	ASLA_ECOLI
22	47	31.3	1030	1	TRI_THEVO
23	46	30.7	240	1	DAPB_STAAM
24	46	30.7	240	1	DAPB_STAAN
25	46	30.7	240	1	DAPB_STAAN
26	46	30.7	716	1	IF2_BACSU
27	45.5	30.3	517	1	GUAA_VIBPA
28	45	30.0	449	1	CMGA_BOVIN
29	45	30.0	457	1	CMGA_HUMAN
30	45	30.0	548	1	GIP2_YEAST
31	45	30.0	866	1	AREA_ASPOR
32	45	30.0	1172	1	TRP2_MOUSE
33	44.5	29.7	523	1	GUAA_HAEIN

34 44.5 29.7 523 1 GUAA_PASMU
35 44.5 29.7 525 1 GUAA_ECOLI6
36 44.5 29.7 525 1 GUAA_ECOLI
37 44.5 29.7 525 1 GUAA_SALTI
38 44 29.3 151 1 LCT2_BOVIN
39 44 29.3 215 1 TCPL_HUMAN
40 44 29.3 359 1 RS2_SPICI
41 44 29.3 375 1 SR55_DROME
42 44 29.3 376 1 VNS2_EHDV2
43 44 29.3 440 1 YD49_YEAST
44 44 29.3 454 1 NCAP_CVM3
45 44 29.3 454 1 NCAP_CVMA5

ALIGNMENTS

RESULT 1
POLG_HCVJ2
ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11111;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE-92230232; PubMed-1314459;
RX Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.*;
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN M AND MRNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: D10074; BAA00968.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
KW INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383

CELLULAR AMINOPEPTIDASE.
CAPSID PROTEIN C (POTENTIAL).
MATRIX PROTEIN (POTENTIAL).
MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 >513 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 513 AA: 55704 MW: 943F31E3514CDEF3 CRC64;

Query Match 92.0%; Score 138; DB 1; Length 513;
 Best Local Similarity 89.3%; Pred. No. 8e-13;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 2

POLG_HCVH4 STANDARD; PRT; 520 AA.
 ID POLG_HCVH4
 AC Q01404;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 DE Hepatitis C virus [Isolate HCV-476] (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID-31643;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93019030; PubMed-1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; D10688; BAA01530.1;
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 >520 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CARBOHYD 196 196 BY SIMILARITY.
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 520 AA: 56499 MW: AAL35246CF20D525 CRC64;

Query Match 92.0%; Score 138; DB 1; Length 520;
 Best Local Similarity 89.3%; Pred. No. 8.1e-13;

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 3

POLG_HCVH4 STANDARD; PRT; 520 AA.
 ID POLG_HCVH4
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 DE Hepatitis C virus [Isolate HCV-KF] (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID-31644;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93019030; PubMed-1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 of non-A, non-B hepatitis in Japan.";
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; D10687; BAA01529.1;
 CC PIR; J01925; J01925.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.

Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NSI; 1.
 ProDom: PD186062; HCV_NSI; 1.
 Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 Transmembrane; Nonstructural protein.
 TRANSMEMBRANE
 INIT_MET 1 1
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 733
 CHAIN 734 >737
 CHAIN 347 369
 TRANSMEM 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 477 477
 CARBOHYD 534 534
 CARBOHYD 542 542
 CARBOHYD 558 558
 CARBOHYD 578 578
 CARBOHYD 627 627
 CARBOHYD 649 649
 NON_TER 737 737
 SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;
 Query Match 92.0%; Score 138; DB 1; Length 737;
 Best Local Similarity 89.3%; Pred. No. 1.2e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQRTKNTNRPPQDVRFSGGGQIVG 28
 DQ 1 PKPQRTKNTNRPPQDVRFSGGGQIVG 28
 DQ 5 PKPQRTKNTNRPPQDVRFSGGGQIVG 32
 DQ 5 PKPQRTKNTNRPPQDVRFSGGGQIVG 32
 RESULT 5
 POLG_HCVJ7
 ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC NCBI_TaxID=11114;
 RN [1]
 RN MEDLINE=92230232; PubMed=1314459;
 RX Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RA "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes".
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC This SWISS-PROT entry is copyright It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D10077; BAA00971.1; -
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 92.0%; Score 138; DB 1; Length 737;
 Best Local Similarity 89.3%; Pred. No. 1.2e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 PKPQKTERNTNRPPQDVRFSGGGQIVG 28
 1 PKPQKTERNTNRPPQDVRFSGGGQIVG 28
 5 PKPQKTERNTNRPPQDVRFSGGGQIVG 32

RESULT 6
 POLG_HCVBK
 ID POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2
 DE (GP68) (NS1); Protein p7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
 DE Hepatitis C virus (isolate BK) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11105;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91140698; PubMed-1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE-96235224; PubMed-8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase.";
 RL Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE-97015086; PubMed-8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moosaw E.W., Adachi T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 RL Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE-98227846; PubMed-9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci. 7:837-847(1998).
 CC [1]- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC [1]- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC [1]- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC [1]- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC [1]- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M58335; AAA72945.1; -
 DR PIR; A38465; GNMVTC.
 DR PDB; 1A1Q; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QUV; 26-JUN-00.
 DR PDB; 80HM; 20-APR-99.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.

DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: SM0186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.
 INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN 3011 369
 TRANSMEM 347 1083
 ACT_SITE 1107 1107
 ACT_SITE 1165 1165
 NP_BIND 1230 1237
 SITE 1316 1319
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 250 250
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 532 532
 CARBOHYD 540 540
 CARBOHYD 556 556
 CARBOHYD 576 576
 CARBOHYD 623 623
 CARBOHYD 645 645
 CARBOHYD 2041 2041
 CARBOHYD 2077 2077
 CARBOHYD 2240 2240
 CARBOHYD 2529 2529
 CARBOHYD 2788 2788
 STRAND 1031 1035
 HELIX 1039 1047
 STRAND 1050 1050
 STRAND 1059 1063
 STRAND 1068 1074
 TURN 1075 1076
 STRAND 1077 1081
 HELIX 1082 1085
 TURN 1086 1087
 STRAND 1090 1092
 TURN 1093 1094
 STRAND 1095 1097
 STRAND 1101 1103
 TURN 1104 1107

InterPro: IPR000745; HCV_NS4a.
 InterPro: IPR001490; HCV_NS4b.
 InterPro: IPR002868; HCV_NS5a.
 InterPro: IPR002166; HCV_RdRP.
 InterPro: IPR007095; RNA_pol_DS_PS.
 InterPro: IPR007094; RNA_pol_PSVir.
 Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.
 Pfam: PF01506; HCV_NS5a; 1.
 Pfam: PF00998; Viral_RdRP; 1.
 ProDom: SM0186062; HCV_NS1; 1.
 SMART: SM00487; DEXDC; 1.
 Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.
 INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN 3011 369
 TRANSMEM 347 1083
 ACT_SITE 1107 1107
 ACT_SITE 1165 1165
 NP_BIND 1230 1237
 SITE 1316 1319
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 250 250
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 532 532
 CARBOHYD 540 540
 CARBOHYD 556 556
 CARBOHYD 576 576
 CARBOHYD 623 623
 CARBOHYD 645 645
 CARBOHYD 2041 2041
 CARBOHYD 2077 2077
 CARBOHYD 2240 2240
 CARBOHYD 2529 2529
 CARBOHYD 2788 2788
 STRAND 1031 1035
 HELIX 1039 1047
 STRAND 1050 1050
 STRAND 1059 1063
 STRAND 1068 1074
 TURN 1075 1076
 STRAND 1077 1081
 HELIX 1082 1085
 TURN 1086 1087
 STRAND 1090 1092
 TURN 1093 1094
 STRAND 1095 1097
 STRAND 1101 1103
 TURN 1104 1107

FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDF9C CRC64;
 Query Match 92.0%; Score 138; DB 1; Length 3010;
 Best Local Similarity 89.3%; Pred. No. 5.5e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQRTKTRNTNRPPQVRFSGGGQIVG 28
 DB 5 PKPQRTKTRNTNRPPQVRFSGGGQIVG 32
 RESULT 7
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).
 OS Hepatitis C virus [isolate Japanese] (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 Sugimura T., Shimotohno K.;
 RA "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

FT	CARBOHYD	196	196	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	209	209	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	234	234	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	305	305	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	417	417	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	423	423	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	430	430	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	448	448	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	476	476	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	532	532	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	540	540	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	556	556	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	576	576	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	623	623	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	CARBOHYD	645	645	N-LINKED (GLCNAC.	.)	(POTENTIAL)
FT	STRAND	1224	1226				
FT	TURN	1232	1233				
FT	TURN	1236	1238				
FT	HELIX	1239	1246				
FT	TURN	1247	1248				
FT	STRAND	1251	1255				
FT	HELIX	1258	1271				
FT	TURN	1272	1272				
FT	STRAND	1277	1280				
FT	TURN	1281	1282				
FT	STRAND	1283	1285				
FT	STRAND	1291	1295				
FT	HELIX	1296	1301				
FT	TURN	1302	1303				
FT	STRAND	1312	1316				
FT	TURN	1317	1319				
FT	HELIX	1323	1335				
FT	TURN	1336	1340				
FT	STRAND	1343	1347				
FT	TURN	1352	1353				
FT	TURN	1360	1361				
FT	STRAND	1362	1366				
FT	STRAND	1368	1368				
FT	STRAND	1373	1375				
FT	TURN	1376	1377				
FT	STRAND	1378	1380				
FT	HELIX	1382	1385				
FT	STRAND	1389	1393				
FT	HELIX	1397	1409				
FT	TURN	1410	1411				
FT	STRAND	1414	1417				
FT	TURN	1419	1420				
FT	STRAND	1432	1436				
FT	TURN	1438	1439				
FT	STRAND	1450	1453				
FT	STRAND	1456	1463				
FT	STRAND	1471	1478				
FT	STRAND	1480	1480				
FT	HELIX	1481	1488				
FT	TURN	1489	1490				
FT	STRAND	1497	1501				
FT	STRAND	1507	1507				
FT	STRAND	1511	1511				
FT	HELIX	1514	1527				
FT	HELIX	1532	1544				
FT	STRAND	1550	1550				
FT	HELIX	1555	1564				
FT	HELIX	1570	1578				
FT	TURN	1579	1580				
FT	HELIX	1584	1597				
FT	TURN	1598	1598				
FT	HELIX	1606	1611				
FT	TURN	1614	1618				
FT	STRAND	1622	1623				
FT	STRAND	1627	1627				
FT	STRAND	1635	1636				
FT	HELIX	1640	1652				
SO	SEQUENCE	3011	AA: 327142 MW: 772CBB29CCD94753 CRC64:				

FT	HELIX
SQ	SEQUENCE

Query Match 92.0%; Score 138; DB 1; Length 3011;
 Best Local Similarity 89.3%; Pred. No. 5.5e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
 DB 5 PKPQKTRNTNRPPQDVRFSGGGQIVG 32

RESULT 9
 POLG_HCVJ6 STANDARD; PRT: 3033 AA.
 ID POLG_HCVJ6
 AC P26660;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP33) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A. PubMed=1658196;
 RX MEDLINE=9204440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions".
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; D00944; BAA00792.1; -
 DR PIR; J01303; J01303.
 DR HSSP; P27958; 1HEI.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4A.
 DR InterPro; IPR001490; HCV_NS4B.

DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_FSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural
 KW REMOVED FROM CAPSID PROTEIN C BY THE
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 92.0%; Score 138; DB 1; Length 3033;
 Best Local Similarity 89.3%; Pred. No. 5.6e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
 DB 5 PKPQKTRNTNRPPQDVRFSGGGQIVG 32

RESULT 10
 POLG_HCVJ8 STANDARD; PRT: 3033 AA.
 ID POLG_HCVJ8
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT *Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes*.
 Virology 186:331-341(1992).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL: D10988; BAA01761.1; -;
 DR PIR: A40250; GNMVJ8.
 DR HSSP: P27958; 1HEI.
 DR MEROPS: S29.001; -;
 DR MEROPS: U39.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PsVIR.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2359 2359
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E73381FD1A CRC64;

Query Match 92.0%; Score 138; DB 1; Length 3033;
 Best Local Similarity 89.3%; Pred. No. 5.6e-12;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRPPQDVRFSGGQIVG 28
 |||||:|||||:|||||:|||||
 DB 5 PKPQKTRNTNRPPQDVRFSGGQIVG 32

RESULT 11

POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: capsid protein c (core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA.;
 Virology 188:102-113(1992).
 -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the p6 position, Cys or Thr in p1 and Ser or Ala in p1'.
 -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL: M84754; -- NOT_ANNOTATED_CDS.
 DR PIR: A40244; GNMVWTW.
 DR PDB: 1N64; 25-FEB-03.
 DR MOP: 1NS3; 08-APR-98.
 DR MEROPS: S29.001; --
 DR MEROPS: U39.001; --
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PsVsr.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 3010 347 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH_BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2329 2329 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 87.3%; Score 131; DB 1; Length 3010;
 Best Local Similarity 88.9%; Pred. No. 6e-11;
 Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 KPQRKTRNTNRPPQDVRFSGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 6 KPQRKTRNTNRPPQDVRFSGGQIVG 32
 RESULT 12
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC 000269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE Hepatitis C virus (isolate HC-JT) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; Pubmed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals*";
 RL Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M84754; -- NOT_ANNOTATED_CDS.
 DR PIR: A40244; GNMVWTW.
 DR PDB: 1N64; 25-FEB-03.
 DR MOP: 1NS3; 08-APR-98.
 DR MEROPS: S29.001; --
 DR MEROPS: U39.001; --
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PsVsr.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT CHAIN 3010 347 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347

LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D11168; BAA01943.1; -
 PIR: A45573; A45573.
 PDB: 1AIQ; 25-MAR-98.
 PDB: 1JXP; 14-JAN-98.
 MEROPS: S29.001; -
 MEROPS: U39.001; -

InterPro: IPR001410; DEAD.
 InterPro: IPR002522; HCV_capsid.
 InterPro: IPR002521; HCV_core.
 InterPro: IPR002519; HCV_env.
 InterPro: IPR002531; HCV_NS1.
 InterPro: IPR002518; HCV_NS2.
 InterPro: IPR004109; HCV_NS3.
 InterPro: IPR000745; HCV_NS4b.
 InterPro: IPR001490; HCV_NS4d.
 InterPro: IPR002868; HCV_NS5a.
 InterPro: IPR002166; HCV_RDRP.
 InterPro: IPR007095; RNA_pol_DS_PS.
 InterPro: IPR007094; RNA_pol_PSVir.
 Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.
 Pfam: PF01506; HCV_NS5a; 1.
 Pfam: PF00271; helicase_C; 1.
 Pfam: PF00998; Viral_RDRP; 1.
 ProDom: PD186062; HCV_NS1; 1.
 SMART: SM00487; DRXD; 1.

Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.

INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 CHAIN 1616 1862 NON-STRUCTURAL PROTEIN NS4A (POTENTIAL).
 CHAIN 1863 2013 NON-STRUCTURAL PROTEIN NS4B (POTENTIAL).
 CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 CHAIN 3011 369 POTENTIAL.
 CHAIN 370 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CHAIN 1084 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CHAIN 1108 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CHAIN 1166 1237 ATP (POTENTIAL).
 CHAIN 1238 1319 DECH_BOX.
 CHAIN 1320 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CHAIN 197 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CHAIN 210 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CHAIN 235 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CHAIN 251 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CHAIN 306 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CHAIN 418 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D6428B CRC64;

Query Match 86.7%; Score 130; DB 1; Length 3010;
 Best Local Similarity 85.7%; Pred. No. 8.5e-11;
 Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKRTNTNRPPQDVRFSGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKNTYRRPPQDVRFSGGQIVG 32

RESULT 13
 POLG_HCV1
 ID POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OC NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo O.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: M62321; AAA45676.1; -

DR PIR; A39166; GNMVC3.
 DR PDB; 1A1V; 16-FEB-99.
 DR MEROPS; LHEI; 25-NOV-98.
 DR MEROPS; S29 001; -.
 DR MEROPS; U39 001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RDRP.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR007095; RNA_pol_US_PS.
 DR InterPro; IPR007094; RNA_pol_Psvir.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01539; HCV_core; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RDRP; 1.
 DR Prodom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 SO SEQUENCE 3011 AA; 327197 MW; 65P8C9447FCE5AF9 CRC64;
 Query Match 86.7%; Score 130; DB 1; Length 3011;
 Best Local Similarity 82.1%; Pred. No. 8.5e-11;
 Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKTERNTNRPPQDVRFSGGGQIVG 28
 Db 5 PKPQKKNRNTNRPPQDVRFSGGGQIVG 32
 RESULT 14
 2A5D_HUMAN
 ID 2A5D_HUMAN STANDARD; PRT; 602 AA.
 AC Q14738; O00494; O00696; Q15171;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Serine/threonine protein phosphatase 2A, 56 kDa regulatory subunit,
 DE delta isoform (PP2A, B subunit, B' delta isoform) (PP2A, B subunit,
 DE B56 delta isoform) (PP2A, B subunit, PR61 delta isoform) (PP2A, B
 DE subunit, R5 delta isoform).
 GN PP2RSD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM DELTA-1).
 RC TISSUE=Fetal brain;
 RX MEDLINE=96355607; PubMed=8703017;
 RA McGrath B., Rivers A.M., Audlin S., Virshup D.M.;
 RT "The B56 family of protein phosphatase 2A (PP2A) regulatory subunits
 RT encodes differentiation-induced phosphoproteins that target PP2A to
 RT both nucleus and cytoplasm.";
 RL J. Biol. Chem. 271:22081-22089(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS DELTA-1 AND DELTA-3).
 RC TISSUE=Brain cortex;
 RX MEDLINE=97324098; PubMed=9180267;
 RA Tanabe O., Gomez G.A., Nishito Y., Usui H., Takeda M.;
 RT "Molecular heterogeneity of the cDNA encoding a 74-kDa regulatory
 RT subunit (B' or delta) of human protein phosphatase 2A.";
 RL FEBS Lett. 408:52-56(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM DELTA-2), AND SEQUENCE OF 501-508;
 RP 550-559; 573-580 AND 584-601 (DELTA-1).
 RC TISSUE=Bone marrow, and Brain cortex;
 RX MEDLINE=96159032; PubMed=8566219;
 RA Tanabe O., Nagase T., Murakami T., Nozaki H., Usui H., Nishito Y.,
 RA Hayashi H., Kagamiyama H., Takeda M.;
 RT "Molecular cloning of a 74-kDa regulatory subunit (B' or delta) of
 RT human protein phosphatase 2A.";
 RL FEBS Lett. 379:107-111(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS DELTA-1 AND DELTA-2).
 RC TISSUE=Colon, Eye, and Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

Matches 12; Conservative 11; Mismatches 4; Indels 11; Gaps 2

QY 1 PKPQRKTE-RNTNRRRPD-----VRFSGGGQIV 27
I:I :::: :I:I I:
DB 44 PQQAQSQPSSNKRPSTPTPTQTISKIKYSGPQIV 81

RESULT 15
2A5D_RABIT
ID 2A5D_RABIT STANDARD; PRT; 586 AA.
AC Q28653; Q28653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 39, Last sequence update)
DE Serine/threonine protein phosphatase 2A, 56 kDa regulatory subunit,
DE delta isoform (PP2A, B subunit, B' delta isoform) (PP2A, B subunit,
DE B56 delta isoform) (PP2A, B subunit, PR61 delta isoform) (PP2A, B
DE subunit, R5 delta isoform) (PP2A, B subunit, B'-gamma).
GN PP2R5D.

OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-New Zealand; TISSUE-Brain, and Skeletal muscle;
RX MEDLINE=9616194; PubMed=8576224;
RA Scioros C., Zolnierowicz S., Bakto E., Durbin S.D., Depaoli-Roach A.A.;
RT "High complexity in the expression of the B' subunit of protein
RT phosphatase 2A0. Evidence for the existence of at least seven novel
RT isoforms.*";
RL J. Biol. Chem. 271:2578-2588(1996).
CC -I- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC COMPARTMENT.
CC -I- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES
CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
CC B (THE R2/B/PR55/R55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
CC FAMILIES), THE 48 KDA VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC AND CELL SIGNALING MOLECULES.
CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -I- TISSUE SPECIFICITY: Highly expressed in brain.
CC -I- SIMILARITY: Belongs to the phosphatase 2A regulatory subunit B
CC family.
CC CAUTION: NOMENCLATURE USED IN REF.1 REFERS TO PP2A B SUBUNIT
CC B' GAMMA ISOFORM, WHICH IS CITED AS PP2A B SUBUNIT DELTA-PR61
CC ISOFORM IN LATER PUBLICATIONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL; U38193; AAC48532.1; -;
DR EMBL; U38195; AAC48534.1; -;
DR InterPro: IPR002554; B56.
DR Pfam: PF01603; B56; 1
KW Multigene family; Nuclear protein; Repeat.
FT DOMAIN 21 36 8 X 2 AA APPROXIMATE TANDEM REPEATS OF Q-
FT P.
FT DOMAIN 507 514 SH3 BINDING, CLASS I (POTENTIAL).
FT DOMAIN 532 549 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 586 AA; 68090 MW; E149A309CDDA7435 CRC64;

Query Match
Best Local Similarity 31.6%; DB 1; Length 586;
Pred. No. 4.3;

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	141	94.0	103	12	Q9E977	Q9E977	hepatitis c
2	141	94.0	128	12	Q81481	Q81481	hepatitis c
3	141	94.0	191	12	Q8V847	Q8V847	hepatitis c
4	139	92.7	191	12	Q68137	Q68137	hepatitis c
5	138	92.0	45	12	Q68311	Q68311	hepatitis c
6	138	92.0	45	12	Q68306	Q68306	hepatitis c
7	138	92.0	46	12	Q68309	Q68309	hepatitis c
8	138	92.0	60	12	Q8JYR9	Q8JYR9	hepatitis c
9	138	92.0	61	12	Q8JYR8	Q8JYR8	hepatitis c
10	138	92.0	61	12	Q8JYR5	Q8JYR5	hepatitis c
11	138	92.0	61	12	Q8JYR7	Q8JYR7	hepatitis c
12	138	92.0	62	12	Q8JYR6	Q8JYR6	hepatitis c
13	138	92.0	74	12	Q68712	Q68712	hepatitis c
14	138	92.0	74	12	Q68682	Q68682	hepatitis c
15	138	92.0	74	12	Q68708	Q68708	hepatitis c
16	138	92.0	74	12	Q68679	Q68679	hepatitis c


```
Db 2 PKPQKTKRNTNRRPQDVRFPGGGQIVG 29
RESULT 2
Q81481
ID Q81481 PRELIMINARY; PRT; 128 AA.
AC Q81481;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GIUS1;
RA Okamoto H., Crovatto M., Pozzato G., Feray C., Brechot C., Mishirow S.;
"molecular and clinical characteristics of the hepatitis C virus
genotype '2c' found in Italians in Italy and France.";
Int. Hepatol. Commun. 3:161-165(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; D31986; BAA06754.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 1 1
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 14591 MW; 0F5F0083F8ED0B5A CRC64;

Query Match 94.08; Score 141; DB 12; Length 128;
Best Local Similarity 92.98; Pred. No. 3.3e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFPGGGQIVG 28
Db 3 PKPQKTKRNTNRRPQDVRFPGGGQIVG 30
RESULT 3
Q8V847
ID Q8V847 PRELIMINARY; PRT; 191 AA.
Q8V847;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Liver;
RA Kato N.;
RT "Hepatitis C virus quasispecies in cancerous and non-cancerous
lesions.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AB061932; BAB3271.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.

KW Polyprotein.
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20902 MW; 43A1B0CEFB44A0F9 CRC64;

Query Match 94.08; Score 141; DB 12; Length 191;
Best Local Similarity 92.98; Pred. No. 5e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFPGGGQIVG 28
Db 5 PKPQKTKRNTNRRPQDVRFPGGGQIVG 32
RESULT 4
Q68137
ID Q68137 PRELIMINARY; PRT; 191 AA.
AC Q68137;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
CN C.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SA6;
RX MEDLINE-94336721; PubMed-8058787;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the core gene of 14 hepatitis C virus
genotypes.";
Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SA6;
RX MEDLINE-92279243; PubMed-1317578;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the 5' noncoding region of hepatitis C virus.";
Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SA6;
RX MEDLINE-93376778; PubMed-8396266;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "At least 12 genotypes of hepatitis C virus predicted by sequence
analysis of the putative E1 gene of isolates collected worldwide.";
Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; U10220; AAA21059.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20682 MW; 0EB8000CF1F902EB CRC64;

Query Match 92.74; Score 139; DB 12; Length 191;
Best Local Similarity 89.38; Pred. No. 1e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTNRRPQDVRFPGGGQIVG 28
Db 5 PKPQKTKRNTNRRPQDVRFPGGGQIVG 32
RESULT 5
Q68311
```

```

ID Q68311 PRELIMINARY; PRT; 45 AA.
AC Q68311;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB37;
RA Songvilalai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23749; AAA65056.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5015 MW; CC527167096AA81 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 45;
Best Local Similarity 89.3%; Pred. No. 2.9e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTNRRPDVRFSGGGQIVG 28
DB 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 6
ID Q68306 PRELIMINARY; PRT; 45 AA.
AC Q68306;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB5;
RA Songvilalai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23744; AAA65051.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5030 MW; CC4C21ED236AA81 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 45;
Best Local Similarity 89.3%; Pred. No. 2.9e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTNRRPDVRFSGGGQIVG 28
DB 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

```

```

DB 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 7
ID Q68309 PRELIMINARY; PRT; 46 AA.
AC Q68309;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB10;
RA Songvilalai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23747; AAA65054.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 46 AA; 5129 MW; ABDC4C21ED236AAA CRC64;

Query Match 92.0%; Score 138; DB 12; Length 46;
Best Local Similarity 89.3%; Pred. No. 3e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTNRRPDVRFSGGGQIVG 28
DB 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 8
ID Q8JYR9 PRELIMINARY; PRT; 60 AA.
AC Q8JYR9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG306;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; AF506616; AAM33392.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 60 AA; 6586 MW; B5691CBE8F26F2F1 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 60;
Best Local Similarity 89.3%; Pred. No. 4e-13;

```

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 9

Q8JYR8 PRELIMINARY; PRT; 61 AA.
 AC Q8JYR8;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-RIG286;
 RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
 RT "Genetic variability of hepatitis C virus in Western Siberia.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF506617; AAM33393.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 61
 SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 61;
 Best Local Similarity 89.3%; Pred. No. 4.1e-13;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 10

Q8JYR5 PRELIMINARY; PRT; 61 AA.
 Q8JYR5;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-RIG290;
 RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
 RT "Genetic variability of hepatitis C virus in Western Siberia.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF506620; AAM33396.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 61
 SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 61;
 Best Local Similarity 89.3%; Pred. No. 4.1e-13;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 11

Q8JYR7 PRELIMINARY; PRT; 61 AA.
 AC Q8JYR7;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-RIG288;
 RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
 RT "Genetic variability of hepatitis C virus in Western Siberia.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF506618; AAM33394.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 61
 SQ SEQUENCE 61 AA; 6869 MW; EC657F706E8F26F2 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 61;
 Best Local Similarity 89.3%; Pred. No. 4.1e-13;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 12

Q8JYR6 PRELIMINARY; PRT; 62 AA.
 AC Q8JYR6;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-RIG289;
 RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
 RT "Genetic variability of hepatitis C virus in Western Siberia.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF506619; AAM33395.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.

Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 9

Q8JYR8 PRELIMINARY; PRT; 61 AA.
 AC Q8JYR8;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-RIG286;
 RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
 RT "Genetic variability of hepatitis C virus in Western Siberia.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF506617; AAM33393.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 61
 SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match 92.0%; Score 138; DB 12; Length 61;
 Best Local Similarity 89.3%; Pred. No. 4.1e-13;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
 |||||:|||||:|||||:|||||
 Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 10

Q8JYR5 PRELIMINARY; PRT; 61 AA.
 Q8JYR5;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-RIG290;
 RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
 RT "Genetic variability of hepatitis C virus in Western Siberia.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF506620; AAM33396.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 61
 SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

```

KW Polyprotein. 62      62
FT NON_TER      62 AA: 7053 MW: EEE656DC79E9F26 CRC64;
SQ SEQUENCE      62 AA: 7053 MW: EEE656DC79E9F26 CRC64;

Query Match      92.0%; Score 138; DB 12; Length 62;
Best Local Similarity 89.3%; Pred. No. 4.2e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 13
ID Q68712 PRELIMINARY; PRT; 74 AA.
AC Q68712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 1b.
RT "Hepatitis C virus genotyping by means of 5'-UR/core line probe assays
and molecular analysis of untypeable samples.";
RL Virus Res. 38:137-157(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: L38421; AAC42199.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
FT NON_TER
SQ SEQUENCE 74 AA: 8470 MW: 2AE2CF5ED0BD10F3 CRC64;

Query Match      92.0%; Score 138; DB 12; Length 74;
Best Local Similarity 89.3%; Pred. No. 5e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 14
ID Q68682 PRELIMINARY; PRT; 74 AA.
AC Q68682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 1b.
RT "Hepatitis C virus genotyping by means of 5'-UR/core line probe assays
and molecular analysis of untypeable samples.";
RL Virus Res. 38:137-157(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: L38421; AAC42199.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
FT NON_TER
SQ SEQUENCE 74 AA: 8470 MW: 2AE2CF5ED0BD10F3 CRC64;

Query Match      92.0%; Score 138; DB 12; Length 74;
Best Local Similarity 89.3%; Pred. No. 5e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 15
ID Q68708 PRELIMINARY; PRT; 74 AA.
AC Q68708;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 1b.
RT "Hepatitis C virus genotyping by means of 5'-UR/core line probe assays
and molecular analysis of untypeable samples.";
RL Virus Res. 38:137-157(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: L38344; AAC42198.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
FT NON_TER
SQ SEQUENCE 74 AA: 8412 MW: 2AE2DB0F976D10F3 CRC64;

Query Match      92.0%; Score 138; DB 12; Length 74;
Best Local Similarity 89.3%; Pred. No. 5e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

Search completed: August 7, 2003, 11:19:02
Job time : 25.6364 secs

```

```

RT and molecular analysis of untypeable samples.";
RL Virus Res. 38:137-157(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: L38348; AAC42168.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
FT NON_TER
SQ SEQUENCE 74 AA: 8440 MW: 2AE2DB0F95BD10F3 CRC64;

Query Match      92.0%; Score 138; DB 12; Length 74;
Best Local Similarity 89.3%; Pred. No. 5e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 15
ID Q68708 PRELIMINARY; PRT; 74 AA.
AC Q68708;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 1b.
RT "Hepatitis C virus genotyping by means of 5'-UR/core line probe assays
and molecular analysis of untypeable samples.";
RL Virus Res. 38:137-157(1995).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: L38344; AAC42198.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
FT NON_TER
SQ SEQUENCE 74 AA: 8412 MW: 2AE2DB0F976D10F3 CRC64;

Query Match      92.0%; Score 138; DB 12; Length 74;
Best Local Similarity 89.3%; Pred. No. 5e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTERNTNRRPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
Db 5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

Search completed: August 7, 2003, 11:19:02
Job time : 25.6364 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-31

Perfect score: 150

Sequence: 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

al number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/6C_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	150	100.0	28	3	US-08-921-887-31
2	139	92.7	191	2	US-08-290-665A-204
3	139	92.7	191	5	PCT-US95-10398-204
4	138	92.0	34	3	US-08-380-160-6
5	138	92.0	43	4	US-09-020-846-36
6	138	92.0	44	3	US-08-380-160-2
7	138	92.0	44	4	US-09-389-756-1
8	138	92.0	45	3	US-08-380-160-1
9	138	92.0	61	1	US-07-946-054-9
10	138	92.0	61	1	US-08-083-947-23
11	138	92.0	61	1	US-08-530-550-3
12	138	92.0	61	1	US-08-262-037-26
13	138	92.0	61	5	PCT-US93-08638-9
14	138	92.0	61	5	PCT-US94-07088-23
15	138	92.0	61	5	PCT-US95-13660-3
16	138	92.0	74	3	US-08-836-075A-10
17	138	92.0	74	4	US-08-635-886C-198
18	138	92.0	100	4	US-08-635-886C-232
19	138	92.0	108	3	US-08-836-075A-14
20	138	92.0	115	1	US-08-324-977-8
21	138	92.0	115	2	US-08-384-616-8
22	138	92.0	115	2	US-08-904-686A-8
23	138	92.0	115	3	US-09-315-850-8
24	138	92.0	123	2	US-08-501-195-2
25	138	92.0	137	3	US-08-836-075A-46
26	138	92.0	138	3	US-08-836-075A-60
27	138	92.0	140	2	US-08-501-195-1

28 138 92.0 154 3 US-08-854-531-2 Sequence 2, Appli
29 138 92.0 154 5 PCT-US95-13552-2 Sequence 2, Appli
30 138 92.0 190 3 US-07-681-701-16 Sequence 16, Appli
31 138 92.0 190 3 US-08-078-271B-1 Sequence 1, Appli
32 138 92.0 191 2 US-08-290-665A-155 Sequence 155, App
33 138 92.0 191 2 US-08-290-665A-157 Sequence 156, App
34 138 92.0 191 2 US-08-290-665A-153 Sequence 157, App
35 138 92.0 191 2 US-08-290-665A-153 Sequence 158, App
36 138 92.0 191 2 US-08-290-665A-153 Sequence 159, App
37 138 92.0 191 2 US-08-290-665A-160 Sequence 160, App
38 138 92.0 191 2 US-08-290-665A-161 Sequence 161, App
39 138 92.0 191 2 US-08-290-665A-163 Sequence 163, App
40 138 92.0 191 2 US-08-290-665A-164 Sequence 164, App
41 138 92.0 191 2 US-08-290-665A-165 Sequence 165, App
42 138 92.0 191 2 US-08-290-665A-165 Sequence 166, App
43 138 92.0 191 2 US-08-290-665A-167 Sequence 167, App
44 138 92.0 191 2 US-08-290-665A-168 Sequence 168, App
45 138 92.0 191 2 US-08-290-665A-163 Sequence 169, App

ALIGNMENTS

RESULT 1
US-08-921-887-31
; Sequence 31, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAM
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-31

Query Match 100.0% Score 150; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 5e15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||

RESULT 2
US-08-290-665A-204
; Sequence 204, Application US/08290665A
; Patent No. 582852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PORCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: SA6
US-08-290-665A-204

Query Match 92.7%; Score 139; DB 2; Length 191;
Best Local Similarity 89.3%; Pred. No. 1.5e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRPPQDVRFSGGGQIVG 32
|||||:|||||:|||||:|||||

RESULT 3
PCT-US95-10398-204
; Sequence 204, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PORCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: SA6
PCT-US95-10398-204

Query Match 92.7%; Score 139; DB 5; Length 191;
Best Local Similarity 89.3%; Pred. No. 1.5e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
Db 5 PKPQKTRNTNRPPQDVRFSGGGQIVG 32
|||||:|||||:|||||:|||||

RESULT 4
US-08-380-160-6
; Sequence 6, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P. O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/380,160
;; FILING DATE:
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/057,471
;; FILING DATE: 06-MAY-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berridge, William P.
;; REGISTRATION NUMBER: 30,024
;; REFERENCE/DOCKET NUMBER: WPB 28682
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703)836-6400
;; TELEFAX: (703)836-2787
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 34 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;; ORIGINAL SOURCE:
;; ORGANISM: Human Hepatitis C Virus
;; US-08-380-160-6

Query Match 92.08; Score 138; DB 3; Length 34;
Best Local Similarity 89.38; Pred. No. 3.2e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 PKPQKTRNTNRPPQDVRFSGGQIVG 28
Db 4 PKPQKTRNTNRPPQDVRFSGGQIVG 31

RESULT 5
US-09-020-846-36
;; Sequence 36, Application US/09020846
;; Patent No. 6322965
;; GENERAL INFORMATION:
;; APPLICANT: YAMAGUCHI, Kenjiro
;; APPLICANT: KASHIMAKUMA, Tomiko
;; APPLICANT: CHIBA, Yukie
;; APPLICANT: YAGI, Shintaro
;; APPLICANT: HASEGAWA, Akira
;; TITLE OF INVENTION: CHIMERA ANTIGEN PEPTIDE
;; NUMBER OF SEQUENCES: 72
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FOLEY & LARDNER
;; STREET: 3000 K Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/020,846
;; FILING DATE: 09-FEB-1998
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 9-027015
;; FILING DATE: 10-FEB-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 8-024045

;; FILING DATE: 09-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wegner, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 053466/0225
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 672-5300
;; TELEFAX: (202) 672-5399
;; INFORMATION FOR SEQ ID NO: 36:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 43 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-09-020-846-36

Query Match 92.08; Score 138; DB 4; Length 43;
Best Local Similarity 89.38; Pred. No. 4.1e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGQIVG 28
Db 5 PKPQKTRNTNRPPQDVRFSGGQIVG 32

RESULT 6
US-08-380-160-2
;; Sequence 2, Application US/08380160
;; Patent No. 6235284
;; GENERAL INFORMATION:
;; APPLICANT: DALBON, Pascal
;; APPLICANT: JOLIVET, Michel
;; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
;; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
;; TITLE OF INVENTION: FOR DETECTING THE LATTER
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: OLIFF & BERRIDGE
;; STREET: P.O. Box 19928
;; CITY: Alexandria
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22320
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/380,160
;; FILING DATE:
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/057,471
;; FILING DATE: 06-MAY-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berridge, William P.
;; REGISTRATION NUMBER: 30,024
;; REFERENCE/DOCKET NUMBER: WPB 28682
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703)836-6400
;; TELEFAX: (703)836-2787
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 44 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO

```

; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
; STRAIN: H77
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..44
; OTHER INFORMATION: /note= "N-terminal sequence of the
; OTHER INFORMATION: protein of the nucleocapside or CORE protein of
; OTHER INFORMATION: the human hepatitis C virus"
US-08-380-160-2

Query Match          92.0%; Score 138; DB 3; Length 44;
Best Local Similarity 89.3%; Pred. No. 4.2e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
Db 4 PKPQKTRNTNRPPQDVRFSGGGQIVG 31

JLT 7
--09-389-756-1
; Sequence 1, Application US/09389756
; Patent No. 6576240
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/09/389,756
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: PCT/FR98/00442
; EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PPT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; Patent No. 6576240
; TITLE: of Hepatitis virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-09-389-756-1

Query Match          92.0%; Score 138; DB 4; Length 44;
Best Local Similarity 89.3%; Pred. No. 4.2e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
Db 4 PKPQKTRNTNRPPQDVRFSGGGQIVG 31

RESULT 8
US-08-380-160-1
; Sequence 1, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel

```

```

; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,160
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 28682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-6400
; TELEFAX: (703)836-2787
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
; STRAIN: H77
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..45
; OTHER INFORMATION: /note= "N-terminal sequence of the
; OTHER INFORMATION: protein of the nucleocapside or CORE protein of
; OTHER INFORMATION: the human hepatitis C virus"
US-08-380-160-1

Query Match          92.0%; Score 138; DB 3; Length 45;
Best Local Similarity 89.3%; Pred. No. 4.3e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTRNTNRPPQDVRFSGGGQIVG 28
Db 5 PKPQKTRNTNRPPQDVRFSGGGQIVG 32

RESULT 9
US-07-946-054-9
; Sequence 9, Application US/07946054
; Patent No. 5582968
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara H
; TITLE OF INVENTION: NO. 5582968el Branched Hybrid and Cluster
; TITLE OF INVENTION: Peptides Effective in Diagnosing and Detecting No. 558296
; TITLE OF INVENTION: No. 5582968-B Hepatitis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

```


ADDRESSEE: United Biomedical Inc.
STREET: 25 Davids Dr.
CITY: Hauppauge
STATE: New York
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,054
FILING DATE: 15-SEP-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34,045
REFERENCE/DOCKET NUMBER: 2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-273-2828
TELEFAX: 516-273-1717
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-946-054-9

Query Match 92.0%; Score 138; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 6e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPQDVRESGGQIVG 28
|||||:|||||:|||||:|||||
DB 4 PKPQKTRNTNRRPQDVRESGGQIVG 31

RESULT 10
US-08-083-947-23
Sequence 23, Application US/08083947
Patent No. 5639594
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
APPLICANT: Hosein, Barbara
TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
IN DIAGNOSING AND DETECTING NO. 5639594-A, NO. 5639594-B Hepat
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: M. Lisa Wilson
STREET: 25 Davids Drive
CITY: Hauppauge
STATE: NY
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,947
FILING DATE: 19930628
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 946,054
FILING DATE: 15-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34045
REFERENCE/DOCKET NUMBER: 2000Z
TELECOMMUNICATION INFORMATION:

TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-083-947-23

Query Match 92.0%; Score 138; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 6e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPQDVRESGGQIVG 28
|||||:|||||:|||||:|||||
DB 4 PKPQKTRNTNRRPQDVRESGGQIVG 31

RESULT 11
US-08-530-550-3
Sequence 3, Application US/08530550
Patent No. 5736321
GENERAL INFORMATION:
APPLICANT: Hosein, Barbara
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Peptides Effective for Diagnosis and
IN DETECTION OF HEPATITIS C INFECTION
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: M. Lisa Wilson
STREET: 25 Davids Drive
CITY: Hauppauge
STATE: NY
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,550
FILING DATE:

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34,045
REFERENCE/DOCKET NUMBER: 2000Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-530-550-3

Query Match 92.0%; Score 138; DB 1; Length 61;
Best Local Similarity 89.3%; Pred. No. 6e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPQDVRESGGQIVG 28
|||||:|||||:|||||:|||||
DB 4 PKPQKTRNTNRRPQDVRESGGQIVG 31

RESULT 12
US-08-262-037-26
Sequence 26, Application US/08262037
Patent No. 5747239

```

;
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hoeslin
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Unknown
;
; US-08-262-037-26
;
; Query Match          92.0%; Score 138; DB 1; Length 61;
; Best Local Similarity 89.3%; Pred. No. 6e-13; Indels 0; Gaps 0;
; Matches 25; Conservative 2; Mismatches 1;
;
; QY 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
;      |||||:|||||:|||||:|||||
; DB 4 PKPQKTKRNTNRRPQDVRFSGGGQIVG 31
;
; RESULT 13
; PCT-US93-08638-9
; Sequence 9, Application PC/TUS9308638
; GENERAL INFORMATION:
; APPLICANT: United Biomedical Inc.
; TITLE OF INVENTION: Novel Branched Hybrid and Cluster Peptides
; TITLE OF INVENTION: Effective in Diagnosing and Detecting Non-A,
; TITLE OF INVENTION: Non-B Hepatitis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: UNITED BIOMEDICAL INC.
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: New York
; COUNTRY: USA
;
; Query Match          92.0%; Score 138; DB 1; Length 61;
; Best Local Similarity 89.3%; Pred. No. 6e-13; Indels 0; Gaps 0;
; Matches 25; Conservative 2; Mismatches 1;
;
; QY 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
;      |||||:|||||:|||||:|||||
; DB 4 PKPQKTKRNTNRRPQDVRFSGGGQIVG 31
;
; RESULT 14
; PCT-US94-07088-23
; Sequence 23, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hoeslin, Barbara
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07088
; FILING DATE: 22-JUNE-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,947
; FILING DATE: 28-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LIN, MARIA C.H.
; REGISTRATION NUMBER: 29323
; REFERENCE/DOCKET NUMBER: 1151-4101PCI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 23:

```

```

;
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08638
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Lisa Wilson
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 9055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-2828
; TELEFAX: 516-273-1717
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08638-9
;
; Query Match          92.0%; Score 138; DB 5; Length 61;
; Best Local Similarity 89.3%; Pred. No. 6e-13; Indels 0; Gaps 0;
; Matches 25; Conservative 2; Mismatches 1;
;
; QY 1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
;      |||||:|||||:|||||:|||||
; DB 4 PKPQKTKRNTNRRPQDVRFSGGGQIVG 31
;
; RESULT 14
; PCT-US94-07088-23
; Sequence 23, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hoeslin, Barbara
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective In
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07088
; FILING DATE: 22-JUNE-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,947
; FILING DATE: 28-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LIN, MARIA C.H.
; REGISTRATION NUMBER: 29323
; REFERENCE/DOCKET NUMBER: 1151-4101PCI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 23:

```

SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 PCT-US94-07088-23

Query Match 92.0%; Score 138; DB 5; Length 61;
 Best Local Similarity 89.3%; Pred. No. 6e-13;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
 Db 4 PKPQKTRNTNRRPQDVRFSGGGQIVG 31

RESULT 15

PCT-US95-13660-3
 ; Sequence 3, Application PC/TUS9513660
 ; GENERAL INFORMATION:

APPLICANT: Hosein, Barbara
 APPLICANT: Wang, Chang Yi
 TITLE OF INVENTION: Peptides Effective for
 TITLE OF INVENTION: Diagnosis and Detection of Hepatitis C Infection
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Morgan & Finnegan, L.L.P.
 STREET: 345 Park Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version
 SOFTWARE: #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/13660
 FILING DATE: 23 October 1995

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/333,573
 FILING DATE: 01 November 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Maria C.H. Lin
 REGISTRATION NUMBER: 29,323
 REFERENCE/DOCKET NUMBER: 1151-4118PC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)758-4800
 TELEFAX: (212)751-8849

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 PCT-US95-13660-3

Query Match 92.0%; Score 138; DB 5; Length 61;
 Best Local Similarity 89.3%; Pred. No. 6e-13;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNTNRRPQDVRFSGGGQIVG 28
 Db 4 PKPQKTRNTNRRPQDVRFSGGGQIVG 31

Search completed: August 7, 2003, 11:23:53
 Job time : 11.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146A-31

Perfect score: 150

Sequence: 1 PKPDKTERNTNRPPQVRFSGGQIVG 28

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

al number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138	92.0	44	15	US-10-367-677-1
2	138	92.0	74	10	US-09-851-138-10
3	138	92.0	91	9	US-09-758-308-1
4	138	92.0	97	10	US-09-756-875-8
5	138	92.0	103	10	US-09-921-397-77
6	138	92.0	108	10	US-09-851-138-14
7	138	92.0	113	10	US-09-921-397-78
8	138	92.0	137	10	US-09-851-138-46
9	138	92.0	138	10	US-09-851-138-60
10	138	92.0	166	11	US-09-899-046-152
11	138	92.0	166	11	US-09-878-281-152
12	138	92.0	169	11	US-09-899-046-42
13	138	92.0	169	11	US-09-899-046-44
14	138	92.0	169	11	US-09-878-281-42
15	138	92.0	169	11	US-09-878-281-44

16	138	92.0	182	10	US-09-929-955-2
17	138	92.0	182	14	US-10-104-966-2
18	138	92.0	191	11	US-09-194-949-3
19	138	92.0	318	10	US-09-851-138-76
20	138	92.0	319	10	US-09-851-138-12
21	138	92.0	319	10	US-09-851-138-18
22	138	92.0	319	11	US-09-899-046-50
23	138	92.0	319	11	US-09-899-046-52
24	138	92.0	319	11	US-09-899-046-54
25	138	92.0	319	11	US-09-899-046-144
26	138	92.0	319	11	US-09-878-281-50
27	138	92.0	319	11	US-09-878-281-52
28	138	92.0	319	11	US-09-878-281-54
29	138	92.0	319	11	US-09-878-281-144
30	138	92.0	809	10	US-09-973-025-50
31	138	92.0	809	11	US-09-899-046-50
32	138	92.0	809	11	US-09-995-808-50
33	138	92.0	809	11	US-09-995-860-50
34	138	92.0	2894	10	US-09-941-611-23
35	138	92.0	2894	15	US-10-044-995-23
36	138	92.0	2985	15	US-10-259-275-40
37	138	92.0	3011	9	US-09-742-659-4
38	138	92.0	3011	10	US-09-952-572-9
39	138	92.0	3011	10	US-09-929-955-1
40	138	92.0	3011	10	US-09-747-419-20
41	138	92.0	3011	11	US-09-891-894-3
42	138	92.0	3011	14	US-10-104-966-1
43	138	92.0	3011	15	US-10-259-275-20
44	138	92.0	3012	10	US-09-238-076-2
45	138	92.0	3012	11	US-09-995-937-2

ALIGNMENTS

RESULT 1

US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10367,677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; TITLE: of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match 92.0% Score 138; DB 15; Length 44;

Best Local Similarity 89.3%; Pred. No. 1.3e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
Db 4 PKPQKTKRNTNRRPDVRFSGGGQIVG 31

RESULT 2
US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US2002018308A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-851-138-10
Query Match 92.0%; Score 138; DB 10; Length 74;
Best Local Similarity 89.3%; Pred. No. 2.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
Db 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 3
US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US2002090607A1
; GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 14114.034902
; CURRENT APPLICATION NUMBER: US/09/758,308

; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 92.0%; Score 138; DB 9; Length 91;
Best Local Similarity 89.3%; Pred. No. 2.8e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
Db 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 4
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 92.0%; Score 138; DB 10; Length 97;
Best Local Similarity 89.3%; Pred. No. 3e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTERNTNRRPDVRFSGGGQIVG 28
Db 5 PKPQKTKRNTNRRPDVRFSGGGQIVG 32

RESULT 5

```
US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; -09-921-397-77

Query Match          92.0%; Score 138; DB 10; Length 103;
Best Local Similarity 89.3%; Pred. No. 3.2e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRPPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
DB 18 PKPQKTKRNTNRPPQDVRFSGGGQIVG 45
    |||||:|||||:|||||:|||||

RESULT 6
US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESS: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-921-397-78

Query Match          92.0%; Score 138; DB 10; Length 113;
Best Local Similarity 89.3%; Pred. No. 3.5e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTERNTNRPPQDVRFSGGGQIVG 28
    |||||:|||||:|||||:|||||
DB 5 PKPQKTKRNTNRPPQDVRFSGGGQIVG 32
    |||||:|||||:|||||:|||||

RESULT 8
US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESS: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
```

```

; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46

Query Match          92.0%; Score 138; DB 10; Length 137;
Best Local Similarity 89.3%; Pred. No. 4.3e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 9
US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-851-138-60

Query Match          92.0%; Score 138; DB 10; Length 138;
Best Local Similarity 89.3%; Pred. No. 4.3e-12;

; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46

Query Match          92.0%; Score 138; DB 10; Length 137;
Best Local Similarity 89.3%; Pred. No. 4.3e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 10
US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-152

Query Match          92.0%; Score 138; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 5.3e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 PKPQKTKRNTNRRPQDVRFSGGGQIVG 28
|||||:|||||:|||||:|||||
5 PKPQKTKRNTNRRPQDVRFSGGGQIVG 32

RESULT 11
US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-152

Query Match          92.0%; Score 138; DB 11; Length 166;
```

Best Local Similarity 89.3%; Pred. No. 5.3e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTRNTNRRPQDVRFSGGQIVG 28
Db 5 PKPQKTRNTNRRPQDVRFSGGQIVG 32

RESULT 12
US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-42

Query Match 92.0%; Score 138; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 5.4e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTRNTNRRPQDVRFSGGQIVG 28
Db 5 PKPQKTRNTNRRPQDVRFSGGQIVG 32

RESULT 13
US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-44

Query Match 92.0%; Score 138; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 5.4e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTRNTNRRPQDVRFSGGQIVG 28
Db 5 PKPQKTRNTNRRPQDVRFSGGQIVG 32

RESULT 14
US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-42

Query Match 92.0%; Score 138; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 5.4e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTRNTNRRPQDVRFSGGQIVG 28
Db 5 PKPQKTRNTNRRPQDVRFSGGQIVG 32

RESULT 15
US-09-878-281-44
; Sequence 44, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-44

Query Match 92.0%; Score 138; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 5.4e-12;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTERNTNRPDVRFSGGGQIVG 28
Db 5 PKPQKTKRNTNRPDVRFSGGGQIVG 32

Search completed: August 7, 2003, 12:01:14
Job time : 14.3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds
(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146A-30

Perfect score: 151

Sequence: 1 PKPQRTKRTIRPDQVKEPPGGVIYV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

tal number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	100.0	28	AA1980	NC mosaic protein
2	134	88.7	28	AA1981	NC mosaic protein
3	134	88.7	28	AA1982	HCV type 3 capsid
4	134	88.7	191	AA1983	Hepatitis C virus
5	134	88.7	191	AA1984	Hepatitis C virus
6	134	88.7	191	AA1985	Hepatitis C virus
7	134	88.7	191	AA1986	Hepatitis C virus
8	133	88.1	191	AA1987	Hepatitis C virus
9	130	86.1	470	AA1988	Encoded by Hepatit

10	130	86.1	470	14	AA1989	Encoded by Hepatit
11	130	86.1	470	14	AA1990	Encoded by Hepatit
12	129	85.4	28	20	AA1991	NC mosaic protein
13	129	85.4	3010	15	AA1992	Blood transmissibl
14	129	85.4	3010	23	AA1993	HCV-S1 full-length
15	128	84.8	30	16	AA1994	Hepatitis C virus
16	128	84.8	30	16	AA1995	Hepatitis C virus
17	128	84.8	40	16	AA1996	Hepatitis C virus
18	128	84.8	40	16	AA1997	Hepatitis C virus
19	128	84.8	50	16	AA1998	Hepatitis C virus
20	128	84.8	82	13	AA1999	Hepatitis C virus
21	128	84.8	82	13	AA2000	Non-A, Non-B Hepat
22	128	84.8	191	17	AA2001	Non-A, Non-B Hepat
23	127	84.1	29	14	AA2002	Hepatitis C virus
24	127	84.1	30	13	AA2003	Sequence of peptid
25	127	84.1	30	16	AA2004	Fragment of open r
26	127	84.1	31	13	AA2005	Hepatitis C virus
27	127	84.1	36	16	AA2006	HCV peptide BCH-43
28	127	84.1	38	14	AA2007	CN14 fragment of H
29	127	84.1	38	14	AA2008	HCV capsid peptide
30	127	84.1	38	14	AA2009	HCV capsid peptide
31	127	84.1	38	15	AA2010	Non-A, non-B hepat
32	127	84.1	43	19	AA2011	Hepatitis C virus
33	127	84.1	44	19	AA2012	Hepatitis C virus
34	127	84.1	44	20	AA2013	Hepatitis C virus
35	127	84.1	44	21	AA2014	Human hepatitis C
36	127	84.1	45	21	AA2015	Human hepatitis C
37	127	84.1	55	13	AA2016	HCV core-envelope
38	127	84.1	55	13	AA2017	HCV core-envelope
39	127	84.1	55	13	AA2018	HCV core-envelope
40	127	84.1	55	13	AA2019	HCV core-envelope
41	127	84.1	55	13	AA2020	HCV core-envelope
42	127	84.1	57	13	AA2021	HCV core-envelope
43	127	84.1	61	13	AA2022	Non-A, Non-B Hepat
44	127	84.1	61	16	AA2023	Peptide VIIIE base
45	127	84.1	61	17	AA2024	Anti-HCV antibody
						Prototype peptide

ALIGNMENTS

RESULT 1

ID: AAY06680 standard; Protein; 28 AA.

XX AAY06680;

AC AC

DT 17-JUN-1999 (first entry)

DE NC mosaic protein amino acid fragment H.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;

XX restriction endonuclease assisted ligation; vaccination.

OS Hepatitis C virus.

XX WO9910506-A1.

XX 04-MAR-1999.

XX 21-AUG-1998; 98WO-US17385.

XX 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1998-204671/17.

XX New mosaic protein, comprising a plurality of homologous antigenic peptides from different genotypes of a species - useful for detecting hepatitis infection in an individual

XX PS Claim 5; Fig 9; 66pp; English.

CC The invention relates to a mosaic protein, comprising a plurality of

CC homologous antigenic peptides from different genotypes of a species. The

CC antigenic peptides are from nucleocapsid (NC) proteins. A method for

CC synthesizing an artificial gene that encodes the mosaic protein is also

CC provided. The method is designated restriction endonuclease assisted

CC ligation (REAL). The mosaic protein and the artificial mosaic protein

CC are useful for detecting a hepatitis infection in an individual. The

CC mosaic gene and protein is also useful for vaccination against the

CC infection, especially hepatitis C. The method of synthesizing the

CC artificial gene and the resulting mosaic protein improve the sensitivity,

CC spectrum of immunoreactivity, and antigen specificity of enzyme

CC immunoassays. This provides improved detection of hepatitis C virus.

CC Sequences AAY06673-683 represent amino acid sequence of each monomer

CC comprising the NC mosaic protein.

XX SQ Sequence 28 AA;

Query Match 100.0%; Score 151; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. NO. 9,2e-15;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGPGGVYV 28

DB 1 PKPQKTKRNTIRRPQDVKFGPGGVYV 28

RESULT 2

AA06673

ID AAY06673 standard; Protein; 28 AA.

XX AC AAY06673;

XX DT 17-JUN-1999 (first entry)

XX DE NC mosaic protein amino acid fragment A.

XX KW Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;

XX KW restriction endonuclease assisted ligation; vaccination.

XX OS Hepatitis C virus.

XX PN WO9910506-A1.

XX PD 04-MAR-1999.

XX 21-AUG-1998; 98WO-US17385.

XX 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PA Fields HA, Khudakov YE;

XX PI WPI; 1999-204671/17.

XX DR New mosaic protein, comprising a plurality of homologous antigenic

XX PT peptides from different genotypes of a species - useful for

XX PT detecting hepatitis infection in an individual

XX PS Claim 5; Fig 9; 66pp; English.

XX The invention relates to a mosaic protein, comprising a plurality of

XX homologous antigenic peptides from different genotypes of a species. The

XX antigenic peptides are from nucleocapsid (NC) proteins. A method for

XX synthesizing an artificial gene that encodes the mosaic protein is also

XX provided. The method is designated restriction endonuclease assisted

XX ligation (REAL). The mosaic protein and the artificial mosaic protein

XX are useful for detecting a hepatitis infection in an individual. The

XX mosaic gene and protein is also useful for vaccination against

XX infection, especially hepatitis C. The method of synthesizing the

CC artificial gene and the resulting mosaic protein improve the sensitivity,

CC spectrum of immunoreactivity, and antigen specificity of enzyme

CC immunoassays. This provides improved detection of hepatitis C virus.

CC Sequences AAY06673-683 represent amino acid sequence of each monomer

CC comprising the NC mosaic protein.

XX SQ Sequence 28 AA;

Query Match 88.7%; Score 134; DB 20; Length 28;

Best Local Similarity 96.2%; Pred. NO. 2.7e-12;

Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGPGGVYV 26

DB 1 PKPQKTKRNTIRRPQDVKFGPGGVYV 26

RESULT 3

AA071258

ID AAB71258 standard; protein; 189 AA.

XX AC AAB71258;

XX DT 18-NOV-2002 (first entry)

XX DE HCV type 3 capsid protein fragment.

XX KW Capsid protein; attenuated vaccine; virucide; antiviral; anti-inflammatory;

XX KW hepatotropic; yellow fever; Japanese encephalitis; dengue;

XX KW classical swine fever; bovine viral diarrhoea; hepatitis C.

XX OS Hepatitis C virus.

XX PN WO200266621-A1.

XX PD 29-AUG-2002.

XX PF 11-FEB-2002; 2002WO-AT00046.

XX PR 21-FEB-2001; 2001AT-0000272.

XX PA (HEINZ/) HEINZ F X.

XX PA (MAND/) MANDL C.

XX PI Heinz FX, Mandl C;

XX DR WPI; 2002-667064/71.

XX PT Attenuated flavivirus live vaccine, useful for protection against e.g.

XX PT yellow fever, comprises virus with attenuating deletion of amino acids

XX PT from the capsid protein -

XX PS Disclosure; Fig 2; 30pp; German.

XX This invention describes a novel attenuated flavivirus live vaccine

XX comprising a flavivirus mutant that has a deletion of at least 4

XX consecutive amino acids from the capsid protein, provided that the

XX C-terminal hydrophobic region is not affected by the deletion. The

XX vaccine of the invention has virucide, anti-inflammatory and hepatotropic

XX activity. The attenuated vaccine, and similar nucleic acid vaccines that

XX encode the mutated capsid protein, are useful for protection against a

XX wide range of flavivirus diseases, e.g. yellow fever, Japanese

XX encephalitis, dengue, classical swine fever, bovine viral diarrhoea and

XX hepatitis C. The specified deletion: (i) produces a reliably attenuated

XX virus that does not revert to virulence; (ii) is exactly defined and does

XX not effect immune responses to important proteins; and (iii) can not

XX generate a non-natural virus by recombination. The mutant viruses

XX eliminate the need to produce large amounts of infectious/virulent

XX viruses, and can be produced with less expense. The protective response

XX to flavivirus lasts significantly longer than that to killed vaccines.

XX This sequence represents a fragment of the capsid protein from hepatitis

XX C virus (HCV) type 1 described in the disclosure of the invention.

```

SQ Sequence 189 AA;
Query Match      88.7%; Score 134; DB 23; Length 189;
Best Local Similarity 96.2%; Pred. No. 2e-11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
   |||||
Db 4 PKPQKTKRNTIRRPQDVKFFGGGQI 29

RESULT 4
AAR92968
ID AAR92968 standard; Protein; 191 AA.
XX
AC AAR92968;
XX
DT 02-OCT-1996 (first entry)
XX
% Hepatitis C virus isolate HK10 core protein.
A HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN WO9605315-A2.
XX
PD 22-FEB-1996.
XX
PF 15-AUG-1995; 95WO-US10398.
XX
AC AAR92968;
XX
PR 15-AUG-1994; 94US-0290665.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX
DR WPI; 1996-139709/14.
DR N-PSDB; AAT16643.
XX
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
XX
PS Claim 4; Page 207; 340pp; English.
XX
AAR92936-R92987 are HCV core proteins derived from 52 different HCV
isolates. Isolated cDNA sequences are used for the prodn. of primers
useful for detecting the presence of HCV in a sample, the primers
are also useful for HCV genotyping. Proteins encoded by the cDNAs
can be used in vaccines for immunising against HCV infection. The
proteins may also be used to detect antibodies against HCV in serum,
saliva, lymphocytes or other mononuclear cells. The antibodies may
be used in the prevention of HCV infection.
XX
SQ Sequence 191 AA;
Query Match      88.7%; Score 134; DB 17; Length 191;
Best Local Similarity 96.2%; Pred. No. 2e-11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
   |||||
Db 5 PKPQKTKRNTIRRPQDVKFFGGGQI 30

RESULT 5
AAR92969
ID AAR92969 standard; Protein; 191 AA.
XX
AC AAR92969;
XX

SQ Sequence 189 AA;
Query Match      88.7%; Score 134; DB 23; Length 189;
Best Local Similarity 96.2%; Pred. No. 2e-11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
   |||||
Db 4 PKPQKTKRNTIRRPQDVKFFGGGQI 29

RESULT 4
AAR92968
ID AAR92968 standard; Protein; 191 AA.
XX
AC AAR92968;
XX
DT 02-OCT-1996 (first entry)
XX
% Hepatitis C virus isolate S52 core protein.
A HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN WO9605315-A2.
XX
PD 22-FEB-1996.
XX
PF 15-AUG-1995; 95WO-US10398.
XX
AC AAR92968;
XX
PR 15-AUG-1994; 94US-0290665.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX
DR WPI; 1996-139709/14.
DR N-PSDB; AAT16643.
XX
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
XX
PS Claim 4; Page 208; 340pp; English.
XX
AAR92936-R92987 are HCV core proteins derived from 52 different HCV
isolates. Isolated cDNA sequences are used for the prodn. of primers
useful for detecting the presence of HCV in a sample, the primers
are also useful for HCV genotyping. Proteins encoded by the cDNAs
can be used in vaccines for immunising against HCV infection. The
proteins may also be used to detect antibodies against HCV in serum,
saliva, lymphocytes or other mononuclear cells. The antibodies may
be used in the prevention of HCV infection.
XX
SQ Sequence 191 AA;
Query Match      88.7%; Score 134; DB 17; Length 191;
Best Local Similarity 96.2%; Pred. No. 2e-11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
   |||||
Db 5 PKPQKTKRNTIRRPQDVKFFGGGQI 30

RESULT 6
AAR92971
ID AAR92971 standard; Protein; 191 AA.
XX
AC AAR92971;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate DK12 core protein.
XX
KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN WO9605315-A2.
XX
PD 22-FEB-1996.
XX
PF 15-AUG-1995; 95WO-US10398.
XX
PR 15-AUG-1994; 94US-0290665.
XX

```

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.

XX Bukh J, Miller RH, Purcell RH;

XX WPI: 1996-139709/14.
 DR N-PSDB; AAT16645.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 PT
 PS Claim 4; Page 209-210; 340pp; English.
 XX
 XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

SV Sequence 191 AA;

Query Match 88.7%; Score 134; DB 17; Length 191;

Best Local Similarity 96.2%; Pred. No. 2e-11;

Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQRTKNTIRRPQDVKFGGGVI 26

DB 5 PKPQRTKNTIRRPQDVKFGGGQI 30

RESULT 7

AAR96547

ID AAR96547 standard; peptide; 319 AA.

XX AAR96547;

AC AAR96547;

XX 10-MAR-1997 (first entry)

XX Hepatitis C virus types 7c(8a) isolates VN4 amino acids 1-317.

XX Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection.

XX Hepatitis C virus.

OS

XX

Key Location/Qualifiers

Misc-difference 144

/label= Met, Leu

Misc-difference 144..149

/label= Val, Ala, Glu, Gly

Misc-difference 156

/label= Met, Thr, Lys, Arg

Misc-difference 157

/label= Val, Ala, Asp, Gly

Misc-difference 161

/label= Gly

/note= "amino acid in this position is designated x in
 the specification, but codon usage shows that the
 only possible amino acid at this pos. is Gly"

Misc-difference 167

/label= Ser, Arg

Misc-difference 167

/label= Val, Ala, Glu, Gly

Misc-difference 171

/label= Gly

/note= "amino acid in this position is designated x in
 the specification, but codon usage shows that the
 only possible amino acid at this pos. is Gly"

Misc-difference 172

/label= Cys, Arg, Ser, Gly

FT Misc-difference 174
 FT /label= Phe, Leu, Ile, Val
 FT Misc-difference 177
 FT /label= Phe, Leu
 FT Misc-difference 232
 FT /label= Met, Val
 FT Misc-difference 233
 FT /label= Asn, Asp
 XX
 PN W09613590-A2.
 XX
 PD 09-MAY-1996.
 XX
 XX 23-OCT-1995; 95WO-EP04155.
 XX
 PR 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Stuyver L;
 XX
 XX WPI: 1996-251460/25.
 DR N-PSDB; AAT27958.

Hepatitis C virus poly:nucleic acid unique to unidentified sub:type

- used to develop probes and primers for new sub:types and vaccines

to prevent and treat infection

Claim 25; Fig 3; 150pp; English.

The sequences AAR96526-R96578 represent novel sequences isolated from

hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,

4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,

3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'

untranslated region (UR), the Core/E1, NS4 or NS5B regions of the

genome. This sequence represents amino acids 1-317 from the HCV types 7c

and 8a isolates VN4.

The new HCV types were isolated from patients with chronic HCV from the

Benelux countries, France, Cameroon and Vietnam, because of their

aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR

amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were

sequenced either directly or partially and used to classify the new

viruses into (sub)types based on comparison with known sequences.

The sequences were used to generate the peptides AAR96424-R96524. The

sequences can also be used to synthesise probes and primers for the

detection of HCV in a sample. The polypeptides can be used to detect

anti-HCV antibodies, for HCV typing or to prevent HCV infections.

Sequence 319 AA;

Query Match 88.7%; Score 134; DB 17; Length 319;

Best Local Similarity 96.2%; Pred. No. 3.5e-11;

Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQRTKNTIRRPQDVKFGGGVI 26

DB 5 PKPQRTKNTIRRPQDVKFGGGQI 30

RESULT 8

AAR92970

ID AAR92970 standard; Protein; 191 AA.

XX AAR92970;

XX 02-OCT-1996 (first entry)

XX Hepatitis C virus isolate S2 core protein.

XX HCV: E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX hepatitis.

OS Hepatitis C virus.

XX WO9605315-A2.
 PN
 CC 22-FEB-1996.
 PD
 CC 15-AUG-1995; 95WO-US10398.
 PF
 CC 15-AUG-1994; 94US-0290665.
 PR
 CC (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI; 1996-139709/14.
 DR N-PSDB; AAT16644.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 -- used to determine HCV genotype and as vaccines against HCV infection

Claim 4; Page 209; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

SQ Sequence 191 AA;

Query Match 88.1%; Score 133; DB 17; Length 191;
 Best Local Similarity 92.3%; Pred. No. 2.9e-11;
 Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPORKTKRNTIRRPQDVKFPGGGVI 26
 |||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTIRRPQDVKFPGGGQI 30

RESULT 9

AAR34473
 ID: AAR34473 standard; Protein; 470 AA.

AC AAR34473;

30-JUL-1993 (first entry)

DE Encoded by Hepatitis C virus clone JK3-A.

XX HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
 KW polymerase chain reaction; diagnostic method.

OS Hepatitis C virus.

XX JP05068562-A.

PD 23-MAR-1993.

PF 30-MAY-1991; 91JP-0153736.

PR 30-MAY-1991; 91JP-0153736.

PA (SANW) SANWA KAGAKU KENKYUSHO CO.

DR WPI; 1993-130638/16.

DR N-PSDB; AAO40431.

PT DNA and cDNA of hepatitis C virus - useful as probes for
 diagnosing HCV infection

XX Claim 4; Page 26-28; 44pp; Japanese.

CC cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B
 CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR
 CC amplification, including clone JK3-A. Primer/probes derived from the
 CC sequences of these clones can be used in diagnostic assays for HCV.
 CC See AAO40425-Q40439.

SQ Sequence 470 AA;

Query Match 86.1%; Score 130; DB 14; Length 470;
 Best Local Similarity 92.3%; Pred. No. 2e-10;
 Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPORKTKRNTIRRPQDVKFPGGGVI 26
 |||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTIRRPQDVKFPGGGQI 30

RESULT 10

AAR34474
 ID: AAR34474 standard; Protein; 470 AA.

AC AAR34474;

DT 30-JUL-1993 (first entry)

DE Encoded by Hepatitis C virus clone JK3-B.

XX HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;
 KW polymerase chain reaction; diagnostic method.

OS Hepatitis C virus.

PN JP05068562-A.

PD 23-MAR-1993.

PF 30-MAY-1991; 91JP-0153736.

PR 30-MAY-1991; 91JP-0153736.

PA (SANW) SANWA KAGAKU KENKYUSHO CO.

DR WPI; 1993-130638/16.

DR N-PSDB; AAO40432.

PT DNA and cDNA of hepatitis C virus - useful as probes for
 diagnosing HCV infection

PS Claim 4; Page 28-30; 44pp; Japanese.

CC cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B
 CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR
 CC amplification, including clone JK3-B. Primer/probes derived from the
 CC sequences of these clones can be used in diagnostic assays for HCV.
 CC See AAO40425-Q40439.

SQ Sequence 470 AA;

Query Match 86.1%; Score 130; DB 14; Length 470;
 Best Local Similarity 92.3%; Pred. No. 2e-10;
 Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPORKTKRNTIRRPQDVKFPGGGVI 26
 |||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTIRRPQDVKFPGGGQI 30

RESULT 11

AAR34475
 ID: AAR34475 standard; Protein; 470 AA.

```

XX AC AAR34475;
XX DT 30-JUL-1993 (first entry)
XX DE Encoded by Hepatitis C virus clone JK3-C.
XX KW HCV; non-A, non-B hepatitis virus; NANBH; liver disease;
XX KW polymerase chain reaction; diagnostic method.
XX OS Hepatitis C virus.
XX PN JF05068562-A.
XX PD 23-MAR-1993.
XX PF 30-MAY-1991; 91JP-0153736.
XX PR 30-MAY-1991; 91JP-0153736.
XX PA (SANW ) SANWA KAGAKU KENKYUSHO CO.
XX WPI: 1993-130638/16.
XX N-PSDB; AA040433.
XX PT DNA and cDNA of hepatitis C virus - useful as probes for
XX PT diagnosing HCV infection
XX PS Claim 4; Page 30-32; 44pp; Japanese.
XX CC cDNA was prepared from HCV genomic RNA. Full-length clone JK1-B
XX CC (9405 nucleotides long) and 14 shorter clones were isolated by PCR
XX CC amplification, including clone JK3-C. Primer/probes derived from the
XX CC sequences of these clones can be used in diagnostic assays for HCV.
XX CC See AA040425-Q40439.
XX SQ Sequence 470 AA;
      Query Match 86.1%; Score 130; DB 14; Length 470;
      Best Local Similarity 92.3%; Pred. NO. 2e-10;
      Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPFGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFPFGGVI 30

RESULT 12
AAY06675
ID AAY06675 standard; Protein; 28 AA.
AAY06675;
... 17-JUN-1999 (first entry)
XX DE NC mosaic protein amino acid fragment C.
XX KW Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
XX KW restriction endonuclease assisted ligation; vaccination.
XX OS Hepatitis C virus.
XX PN WO9910506-A1.
XX PD 04-MAR-1999.
XX PF 21-AUG-1998; 98WO-US17385.
XX PR 25-AUG-1997; 97US-0921887.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Fields HA, Khudiyakov YE;

```

```

XX WPI: 1999-204671/17.
XX DE New mosaic protein, comprising a plurality of homologous antigenic
XX DE peptides from different genotypes of a species - useful for
XX DE detecting hepatitis infection in an individual
XX PS Claim 5; Fig 9; 66pp; English.
XX CC The invention relates to a mosaic protein, comprising a plurality of
XX CC homologous antigenic peptides from different genotypes of a species. The
XX CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
XX CC synthesizing an artificial gene that encodes the mosaic protein is also
XX CC provided. The method is designated restriction endonuclease assisted
XX CC ligation (REAL). The mosaic protein and the artificial mosaic protein
XX CC are useful for detecting a hepatitis infection in an individual. The
XX CC mosaic gene and protein is also useful for vaccination against the
XX CC infection, especially hepatitis C. The method of synthesizing the
XX CC artificial gene and the resulting mosaic protein improve the sensitivity,
XX CC spectrum of immunoreactivity, and antigen specificity of enzyme
XX CC immunoassays. This provides improved detection of hepatitis C virus.
XX CC Sequences AAY06673-683 represent amino acid sequence of each monomer
XX CC comprising the NC mosaic protein.
XX SQ Sequence 28 AA;
      Query Match 85.4%; Score 129; DB 20; Length 28;
      Best Local Similarity 92.3%; Pred. NO. 1.4e-11;
      Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPFGGVI 26
DB 1 PKPQKTKRNTIRRPQDVKFPFGGVI 26

RESULT 13
AAR53417
ID AAR53417 standard; Protein; 3010 AA.
XX AC AAR53417;
XX DT 17-JAN-1995 (first entry)
XX DE Blood transmissible NANBH protein.
XX KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
XX KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
XX KW C100 antibody; HCV RNA; NS5 region.
XX OS Non-A, non-B hepatitis virus.
XX PH Key Location/Qualifiers
XX FT Misc-difference 222 /label= His, Arg
XX FT Misc-difference 226 /label= Cys, Arg
XX FT Misc-difference 246 /label= Leu, Phe
XX FT Misc-difference 263 /label= Asp, Asn
XX FT Misc-difference 291 /label= Phe, Ser
XX FT Misc-difference 311 /label= Gly, Asp
XX FT Misc-difference 398 /label= Ser, Arg, Gly
XX FT Misc-difference 400 /label= Thr, Ala
XX FT Misc-difference 405 /label= Gln, Pro, Leu
XX FT Misc-difference 410 /label= Lys, Arg
XX FT Misc-difference 418

```

FT /label- Gly, Asp
 FT Misc-difference 430
 FT /label- Asn, Asp
 FT Misc-difference 438
 FT /label- Phe, Leu
 FT Misc-difference 478
 FT /label- Arg, Lys
 FT Misc-difference 759
 FT /label- Leu, Val
 FT Misc-difference 1017
 FT /label- Ser, Asn
 FT Misc-difference 1036
 FT /label- Thr, Ala
 FT Misc-difference 1036
 FT /label- Glu, Asp
 FT Misc-difference 1201
 FT /label- Met, Thr
 FT Misc-difference 1205
 FT /label- Met, Ile
 FT Misc-difference 1255
 FT /label- Asn, Tyr
 FT Misc-difference 1263
 FT /label- Gly, Asp
 FT Misc-difference 1455
 FT /label- Asn, Asp
 FT Misc-difference 1828
 FT /label- Ala, Thr
 FT Misc-difference 1895
 FT /label- Gly, Arg
 FT Misc-difference 1896
 FT /label- Gly, Ile
 FT Misc-difference 2143
 FT /label- Glu, Val
 FT Misc-difference 2144
 FT /label- Asp, Glu
 FT Misc-difference 2462
 FT /label- Cys, Arg
 FT Misc-difference 2486
 FT /label- Val, Met
 FT Misc-difference 2488
 FT /label- Lys, Gln
 FT Misc-difference 2844
 FT /label- Leu, Met
 FT Misc-difference 2862
 FT /label- Leu, Gln
 FT Misc-difference 2917
 FT /label- Arg, Leu
 FT Misc-difference 2968
 FT /label- Ser, Gly
 FT Misc-difference 2989
 FT /label- Cys, Arg
 FT Misc-difference 2990
 FT /label- Tyr, Cys
 XX JP06105690-A.
 PD 19-APR-1994.
 XX
 XX 10-MAR-1992; 92JP-0051885.
 XX
 PR 10-MAR-1992; 92JP-0051885.
 XX
 XX (KAEN/) KAENNO K.
 XX WPI; 1994-163130/20.
 DR N-PSDB; AAQ63499.
 XX
 PT Blood-transmissible non-A non-B hepatitis virus DNA - used for
 PT detection of hepatitis virus
 XX
 PS Claim 1; Page 8-20; 22pp; Japanese.
 XX
 XX This sequence is encoded by the genome of a blood transmissible non-A,

CC non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the
 CC primers given in AA063500-35. The amplified fragments are used in the
 CC detection of hepatitis virus. The target DNA was isolated from serum
 CC of chronically infected NANBH patients who were C100 antibody-positive
 CC and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR
 CC were performed on cDNA and the total human NANBH DNA was constructed
 CC from 23 clones.
 XX
 SQ Sequence 3010 AA;
 Query Match 85.4%; Score 129; DB 15; Length 3010;
 Best Local Similarity 92.3%; Pred. No. 2e-09; 2; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 2;
 QY 1 PKPQKTKRNTIRRPQDVKEGGVI 26
 Db 5 PKPQKTKRNTIRRPQDVKEGGVI 30
 RESULT 14
 AAE20477
 ID AAE20477 standard; Protein; 3010 AA.
 XX
 AC AAE20477;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE HCV-S1 full-length polypeptide.
 XX
 KW Nucleic acid construct; expression cassette; non-coding region; NCR;
 KW untranslated region; UTR; anti-viral drug; drug resistance;
 KW HCV-S1; Hepatitis C virus.
 XX
 OS Hepatitis C virus.
 XX
 PN WO200208447-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 20-JUL-2001; 2001WO-IL00669.
 XX
 PR 24-JUL-2000; 2000US-220248P.
 XX
 XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.
 PA (EHLR/) EHLRICH G.
 XX
 PI Tan YH, Lim SP, Lim SG, Hong WJ;
 XX
 DR WPI; 2002-280605/32.
 DR N-PSDB; AAD33038.
 XX
 PT Novel nucleic acid construct useful for detecting the presence of RNA
 PT virus, comprises an expression cassette and a promoter operably linked
 PT to expression cassette for minus strand RNA transcription of the
 PT cassette
 XX
 PS Example 1; Page 70-81; 81pp; English.
 XX
 XX The invention relates to nucleic acid construct which comprises an
 CC expression cassette including a first polynucleotide region including
 CC a 5' non-coding region (NCR) sequence of an RNA virus and at least an
 CC N-terminal portion of a coding sequence of RNA virus, a second
 CC polynucleotide region including a 3' untranslated region (UTR) sequence
 CC of the RNA virus and at least a C-terminal portion of a coding sequence
 CC of the virus and a third polynucleotide region encoding a reporter
 CC molecule, flanked by first and second polynucleotide regions; and a
 CC promoter sequence being operatively linked to expression cassette in a
 CC manner so as to enable a transcription of a minus strand RNA molecule
 CC from the expression cassette. Nucleic acid construct of the invention
 CC is useful for detecting the presence of an RNA virus in a cell. It is
 CC also useful for screening anti-viral drugs and determining drug
 CC resistance of an RNA virus. The present sequence is Hepatitis C virus
 CC (HCV) isolate HCV-S1 full-length polypeptide.


```

XX
SQ Sequence 3010 AA;
    Query Match      85.4%; Score 129; DB 23; Length 3010;
    Best Local Similarity 92.3%; Pred. No. 2e-09; 2; Indels 0; Gaps 0;
    Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNITRRPQDVKFPGGGVI 26
   |||||  |||||  |||||  |||||  |||||
Db 5 PKPQRTKRNITRRPQDVKFPGGGQI 30

RESULT 15
AAR84558
ID AAR84558 standard; peptide: 30 AA.
XX
AC AAR84558;
XX
DT 06-JUN-1996 (first entry)
YV Hepatitis C virus core antigenic peptide.

AW HCV; non-A, non-B hepatitis virus; antigen: immunoassay;
KW interferon treatment; monitoring; antibody titre; detection.
XX
OS Hepatitis C virus.
XX
PN JP07260792-A.
XX
PD 13-OCT-1995.
XX
PF 16-MAR-1994; 94JP-0082160.
XX
PR 16-MAR-1994; 94JP-0082160.
XX
PA (ARIM/) ARIMA T.
PA (TORA ) TORAY IND INC.
XX
XX
DR WPI; 1995-386063/50.
XX
PT Determination of the antibody titre against hepatitis C virus
PT antigen - using dilute solution and HCV antigenic peptide(s)
XX
PS Claim 1; Page 10; 1lpp; Japanese.
XX
CC The present sequence is that of an antigen used in a novel assay for
CC determining anti-HCV antibody titre. A sample is diluted to an
CC absorbance of not more than 2.0 and the peptide antigen is added.
CC The method is useful for monitoring the effect of interferon
    treatment on anti-HCV antibody titres.

SQ Sequence 30 AA;
    Query Match      84.8%; Score 128; DB 16; Length 30;
    Best Local Similarity 92.3%; Pred. No. 2.1e-11;
    Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNITRRPQDVKFPGGGVI 26
   |||||  |||||  |||||  |||||  |||||
Db 5 PKPQRTKRNITRRPQDVKFPGGGQI 30

Search completed: August 7, 2003, 11:14:09
Job time : 39.6364 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-30

Perfect score: 151
Sequence: 1 PKPQRTKRNTRRPQDVKPPGGVIV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

tal number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	88.7	114	2 S41359	genome polyprotein
2	134	88.7	114	2 S41358	genome polyprotein
3	134	88.7	124	2 S41360	genome polyprotein
4	134	88.7	411	2 PC2061	genome polyprotein
5	134	88.7	492	2 S41288	genome polyprotein
6	130	86.1	123	2 S41361	genome polyprotein
7	130	86.1	782	2 S19875	genome polyprotein
8	129	85.4	108	2 S41356	genome polyprotein
9	129	85.4	411	2 PC2060	genome polyprotein
10	129	85.4	3010	1 A45573	genome polyprotein
11	128	84.8	88	2 S21336	genome polyprotein
12	127	84.1	108	2 S41353	genome polyprotein
13	127	84.1	108	2 S41355	genome polyprotein
14	127	84.1	108	2 S41357	genome polyprotein
15	127	84.1	108	2 S41348	genome polyprotein
16	127	84.1	112	2 S41371	genome polyprotein
17	127	84.1	112	2 S41341	genome polyprotein
18	127	84.1	114	2 S41370	genome polyprotein
19	127	84.1	114	2 S41369	genome polyprotein
20	127	84.1	114	2 S41368	genome polyprotein
21	127	84.1	115	2 S41342	genome polyprotein
22	127	84.1	115	2 S41344	genome polyprotein
23	127	84.1	115	2 S41350	genome polyprotein
24	127	84.1	115	2 S41354	genome polyprotein
25	127	84.1	115	2 S41345	genome polyprotein
26	127	84.1	115	2 S41347	genome polyprotein
27	127	84.1	115	2 S41343	genome polyprotein
28	127	84.1	118	2 S41346	genome polyprotein
29	127	84.1	266	2 PQ0393	genome polyprotein

30 127 84.1 369 2 S21471 genome polyprotein
31 127 84.1 441 2 S12707 genome polyprotein
32 127 84.1 513 2 PC1284 genome polyprotein
33 127 84.1 520 2 PQ1925 polyprotein - hepa
34 127 84.1 523 2 PQ1926 polyprotein - hepa
35 127 84.1 550 2 JH0711 genome polyprotein
36 127 84.1 782 2 S19876 genome polyprotein
37 127 84.1 782 2 S18031 genome polyprotein
38 127 84.1 782 2 S18032 genome polyprotein
39 127 84.1 787 2 PQ0677 hypothetical prote
40 127 84.1 874 2 PQ0883 genome polyprotein
41 127 84.1 874 2 JQ0881 genome polyprotein
42 127 84.1 876 2 PQ2219 polyprotein - hepa
43 127 84.1 3010 1 GNWVTC genome polyprotein
44 127 84.1 3010 1 GNWVCJ genome polyprotein
45 127 84.1 3010 1 S18030 genome polyprotein

ALIGNMENTS

RESULT 1

S41359
genome polyprotein - hepatitis C virus (genotype 3, N2) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 3, N2
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S41359
R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341
A;Accession: S41359
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>
A;Cross-references: EMBL:Z29462; NID:g443886; PIDN:CAA82600.1; PID:g443887
A;Experimental source: genotype 3, N2
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 88.7% Score 134; DB 2; Length 114;
Best Local Similarity 96.2%; Pred. No. 3.9e-12;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRRPQDVKPPGGVIV 26

DB 5 PKPQRTKRNTRRPQDVKPPGGQI 30

RESULT 2

S41358
genome polyprotein - hepatitis C virus (genotype 3, N1) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 3, N1
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C;Accession: S41358
R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341
A;Accession: S41358
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>
A;Cross-references: EMBL:Z29461; NID:g443884; PIDN:CAA82599.1; PID:g443885
A;Experimental source: genotype 3, N1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 88.7% Score 134; DB 2; Length 114;

Best Local Similarity 96.2%; Pred. No. 3.9e-12; Mismatches 0; Gaps 0; Indels 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
 |||||
 Db 5 PKPQKTKRNTIRRPQDVKFPGGGQI 30

RESULT 3
 S41360
 genome polyprotein - hepatitis C virus (genotype 3, N3) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 3, N3
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41360
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Molecule type: genomic RNA
 A:Residues: 1-124 <VAN>
 A:Cross-references: EMBL:Z29463; NID:g443888; PIDN:CAA82601.1; PID:g443889
 A:Experimental source: genotype 3, N3
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-124/Product: core protein #status predicted <MAT>

Query Match 88.7%; Score 134; DB 2; Length 124;
 Best Local Similarity 96.2%; Pred. No. 4.3e-12; Mismatches 0; Gaps 0; Indels 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
 |||||
 Db 5 PKPQKTKRNTIRRPQDVKFPGGGQI 30

RESULT 4
 PC2061
 genome polyprotein N2 - hepatitis C virus
 N:Contains: envelope protein E1; nonstructural protein E2/NS1
 C:Species: hepatitis C virus
 C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
 C:Accession: PC2061
 R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
 Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
 A:Title: Identification of the third major genotype of hepatitis C virus in France.
 A:Reference number: PC2060; MUID:94197744; PMID:8147893
 A:Accession: PC2061
 A:Molecule type: mRNA
 A:Residues: 1-411 <LIJ>
 A:Cross-references: GB:L12355; NID:g410169; PIDN:AAA20155.1; PID:g410170
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;
 F:192-383/Product: envelope protein E1 #status predicted <SPE>
 F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
 F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 88.7%; Score 134; DB 2; Length 411;
 Best Local Similarity 96.2%; Pred. No. 1.4e-11; Mismatches 0; Gaps 0; Indels 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
 |||||
 Db 5 PKPQKTKRNTIRRPQDVKFPGGGQI 30

RESULT 5
 S41288
 genome polyprotein - hepatitis C virus (fragment)
 N:Contains: core protein; envelope protein; NS1 protein
 C:Species: hepatitis C virus

C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41288
 R:Seelig, R.
 submitted to the EMBL Data Library, December 1993
 A:Reference number: S41288
 A:Accession: S41288
 A:Molecule type: genomic RNA
 A:Residues: 1-492 <SEE>
 A:Cross-references: EMBL:X76918
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; nonstructural protein;
 F:1-191/Product: core protein #status predicted <COR>
 F:192-372/Product: envelope protein #status predicted <ENV>
 F:373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match 88.7%; Score 134; DB 2; Length 492;
 Best Local Similarity 96.2%; Pred. No. 1.7e-11; Mismatches 0; Gaps 0; Indels 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
 |||||
 Db 5 PKPQKTKRNTIRRPQDVKFPGGGQI 30

RESULT 6
 S41361
 genome polyprotein - hepatitis C virus (genotype 3, N4) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 3, N4
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41361
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41361
 A:Molecule type: genomic RNA
 A:Residues: 1-123 <VAN>
 A:Cross-references: EMBL:Z29464; NID:g443890; PIDN:CAA82602.1; PID:g443891
 A:Experimental source: genotype 3, N4
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-123/Product: core protein #status predicted <MAT>

Query Match 86.1%; Score 130; DB 2; Length 123;
 Best Local Similarity 88.5%; Pred. No. 1.6e-11; Mismatches 2; Gaps 0; Indels 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
 |||||
 Db 5 PKPQKTKRNTIRRPQDVKFPGGGQI 30

RESULT 7
 S19875
 genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 prote:
 C:Species: hepatitis C virus
 A:Variety: isolate JK3
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
 C:Accession: S19875
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus
 A:Reference number: S18029
 A:Accession: S19875
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
 A:Experimental source: isolate JK3
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructu

F:1-191/Product: core protein #status predicted <MAT1>
 E:192-383/Product: envelope protein 1 #status predicted <MAT2>
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 86.1%; Score 130; DB 2; Length 782;
 Best Local Similarity 92.3%; Pred. No. 1e-10; Indels 0; Gaps 0;
 Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRRRPQDVKPPGGGVI 26
 |||||:|||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTRRRPQDVKPPGGGQI 30

RESULT 8

S41356 genome polyprotein - hepatitis C virus (genotype 2, N5) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N5
 Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 Accession: S41356
 <van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41356
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:Z29459
 A:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <MAT>

Query Match 85.4%; Score 129; DB 2; Length 108;
 Best Local Similarity 92.3%; Pred. No. 2e-11; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRRRPQDVKPPGGGVI 26
 |||||:|||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTRRRPQDVKPPGGGQI 30

RESULT 9

PC2060 genome polyprotein N1 - hepatitis C virus

N:Contains: envelope protein E1; nonstructural protein E2/NS1
 Species: hepatitis C virus
 Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
 Accession: PC2060
 R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
 Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
 A:Title: Identification of the third major genotype of hepatitis C virus in France.
 A:Reference number: PC2060; MUID:94197744; PMID:8147893
 A:Accession: PC2060
 A:Molecule type: mRNA
 A:Residues: 1-411 <LIJ>
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;
 F:192-383/Product: envelope protein E1 #status predicted <SPE>
 F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
 F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 85.4%; Score 129; DB 2; Length 411;
 Best Local Similarity 88.5%; Pred. No. 7.6e-11; Indels 1; Gaps 0;
 Matches 23; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRRRPQDVKPPGGGVI 26
 |||||:|||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTRRRPQDVKPPGGGQI 30

RESULT 10

A45573 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (non)
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hiji
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carr
 A:Reference number: A45573; MUID:92295714; PMID:1318627
 A:Accession: A45573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <VAN>
 A:Cross-references: GB:D11168; GB:D01171; NID:G221612; PIDN:BAA01943.1; PID:G22161
 A:Experimental source: HCV-JT
 A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 85.4%; Score 129; DB 1; Length 3010;
 Best Local Similarity 92.3%; Pred. No. 5.6e-10; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRRRPQDVKPPGGGVI 26
 |||||:|||||:|||||:|||||:|||||:|
 Db 5 PKPQRTKRNTRRRPQDVKPPGGGQI 30

RESULT 11

S21336 genome polyprotein S29 (core protein region) - hepatitis C virus (fragment)
 C:Species: hepatitis C virus
 C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
 C:Accession: S21336
 R:Sato, A.

submitted to the EMBL Data Library, April 1992
 A:Description: A sensitive serodiagnosis of hepatitis C virus infection with two c:
 A:Reference number: S21336
 A:Accession: S21336
 A:Molecule type: genomic RNA
 A:Residues: 1-88 <SAT>
 A:Cross-references: EMBL:X65548; NID:G59492; PIDN:CAA46517.1; PID:G59493
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: polyprotein

Query Match 84.8%; Score 128; DB 2; Length 88;
 Best Local Similarity 92.3%; Pred. No. 2.2e-11; Indels 2; Gaps 0;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRRRPQDVKPPGGGVI 26
 |||||:|||||:|||||:|||||:|||||:|
 Db 11 PKPQRTKRNTRRRPQDVKPPGGGQI 36

RESULT 12

S41353 genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
 N:Contains: core protein

Query Match 84.1%; Score 127; DB 2; Length 108;
Best Local Similarity 92.3%; Pred. No. 3.8e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	PKPQRTKRN	TI	RR	PQDV	K	FGG	VI	26
Db	5	PKPQRTKRN	TI	RR <th>PQDV</th> <th>K</th> <th>FGG</th> <th>QI</th> <th>30</th>	PQDV	K	FGG	QI	30

N: Contains: core protein
C: Species: hepatitis C virus
A: Variety: genotype 1, N6

A:Accession: S41348
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Comments: ENZY. 226457

Query Match	84.1%	Score 127;	DB 2;	Length 108;
Best Local Similarity	92.3%	Pred. No. 3.8e-11;		
Matches 24:	Conservative	0: Mismatches	2: Indels	

Search completed: August 7, 2003, 11:21:48
Job time : 9.54545 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146A-30

Perfect score: 151

Sequence: 1 PKQRTKRTNIRPQDPKPGGVIVV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

cal number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	85.4	3010	1	POLG_HCVJT
2	127	84.1	513	1	POLG_HCVJ2
3	127	84.1	520	1	POLG_HCVH4
4	127	84.1	520	1	POLG_HCVHK
5	127	84.1	737	1	POLG_HCVJ5
6	127	84.1	737	1	POLG_HCVJ7
7	127	84.1	3010	1	POLG_HCVBK
8	127	84.1	3010	1	POLG_HCVJA
9	127	84.1	3011	1	POLG_HCVH
10	127	84.1	3033	1	POLG_HCVJ6
11	127	84.1	3033	1	POLG_HCVJ8
12	120	79.5	3010	1	POLG_HCVTW
13	119	78.8	3011	1	POLG_HCV1
14	53	35.1	282	1	RK4_TOBAC
15	52	34.4	794	1	FURL_HUMAN
16	50.5	33.4	3988	1	POLG_BVDVN
17	49	32.5	293	1	RK4_SPIOI
18	49	32.5	496	1	CAT3_MAIZE
19	49	32.5	1520	1	ACFD_ECOLI
20	48.5	32.1	309	1	FN3X_HUMAN
21	48	31.8	219	1	Y401_BUCAP
22	47	31.1	58	1	RL29_YEAST
23	47	31.1	793	1	FURL_MOUSE
24	46.5	30.8	309	1	FN3K_MOUSE
25	46	30.5	230	1	RL4_MYCLE
26	46	30.5	797	1	FURL_BOVIN
27	45.5	30.1	298	1	YSMK_CAEEL
28	45.5	30.1	303	1	YBL_XENLA
29	45	29.8	223	1	RL4_MYCHO
30	45	29.8	223	1	RL4_MYCTU
31	45	29.8	282	1	RK4_ARATH
32	45	29.8	595	1	FRA_DROME
33	45	29.8	897	1	BGAL_CLOAB

ALIGNMENTS

RESULT 1

POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals."
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D11168; BAA01943.1;
DR PIR; A45573; A45573.
DR PDB; 1A1Q; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR MEROPS; S29.001;
DR MEROPS; U39.001;
DR InterPro; IPR001410; DEAD.

34 45 29.8 941 1 CHRD_XENLA Q91713 xenopus lae
35 44.5 29.5 279 1 Y144_MYCGE P47390 mycoplasma
36 44.5 29.5 732 1 TAU_MOUSE P10637 mus musculus
37 44.5 29.5 751 1 TAU_RAT P19332 rattus norv
38 44.5 29.5 1041 1 SVI_MYCTU Q10765 mycobacteri
39 44 29.1 207 1 RL4_BACSU Q42921 bacillus su
40 44 29.1 332 1 P11_MOUSE Q9CZXS mus musculus
41 44 29.1 512 1 T663_HSV2H P28276 herpes simp
42 44 29.1 775 1 TNP3_MOUSE Q60769 mus musculus
43 44 29.1 858 1 CHS1_RHIOL P30594 rhizopus ol
44 44 29.1 938 1 HDA7_MOUSE Q8C2B3 mus musculus
45 43.5 28.8 206 1 RL31_BRANA P41128 brassica na

```

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_Core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN E (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 NON-STRUCTURAL PROTEIN NS3 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS4 (POTENTIAL).
FT CHAIN 1863 2013 NON-STRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NON-STRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 3011 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 3010 AA; 326573 MW; 94A1C774350642BB CRC64;
Query Match 85.4%; Score 129; DB 1; Length 3010;
Best Local Similarity 92.3%; Pred. No. 1.1e-10;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRTKNTYRRPQDVFPGGQVI 26

```

```

DB 5 PKPQRTKNTYRRPQDVFPGGQI 30
RESULT 2
POLG_HCVJ2 STANDARD; PRT; 513 AA.
ID POLG_HCVJ2
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: capsid protein c (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes. RT
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10074; BAA00968.1; -.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_Core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN E (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).

```

F.T.	CARBOHYD	449	449
------	----------	-----	-----

FI	CARBOHYD	303	N-LINKED	(GLCNAc)
FT	CARBOHYD	418	N-LINKED	(GLCNAc)

TRANSMEM	347	369	POTENTIAL.
FTT	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	477	477	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	534	534	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	542	542	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	558	558	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	578	578	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	649	649	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	737	737	N-LINKED (GLCNAC. . .) (POTENTIAL).
FTT	737	737	N-LINKED (GLCNAC. . .) (POTENTIAL).
NON_TER	737	737	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE	737 AA:	81207 MW; 3AF69D82AD501B1 CRC64;	

Query Match	84.1%	Score 127;	DB 1;	Length 737;
Best Local Similarity	92.3%	Pred. No. 5e-11;		
Matches	24;	Conservative	0;	Mismatches
			2;	Indels
				0;
				Gaps
				0;

QY	1	PKPQKTKRNTIRRPQDVKFPGGGVI	26
DBD	5	PKPQKTKRNTIRRPQDVKFPGGGOI	30

RESULT 6	
POLG_HCVJ7	
ID	POLG_HCVJ7
AC	P27961;
CD	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DE	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Genome polyprotein (Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1 and NS2) (Fragment).
DE	Hepatitis C virus (isolate HC-J7) (HCV).
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepacivirus
OC	NCBI_TaxID=11114;
OC	[1]
NRN	SEQUENCE FROM N.A.
RRP	MEDLINE=92230232; PubMed=1314459;
RRP	Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RRP	Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RRP	"Full-length sequence of a hepatitis C virus genome having poor
RRP	homology to reported isolates: comparative study of four distinct
RRP	genotypes.";
RRP	Virolgy 188:331-341(1992).
RRP	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
RRP	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
RRP	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION
RRP	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
RRP	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
RRP	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
RRP	PROTEIN C AND MRNA.
RRP	-----
RRP	This SWISS-PROT entry is copyright. It is produced through a collaboration
RRP	between the Swiss Institute of Bioinformatics and the EMBL outstation -
RRP	the European Bioinformatics Institute. There are no restrictions on its
RRP	use by non-profit institutions as long as its content is in no way
RRP	modified and this statement is not removed. Usage by and for commercial
RRP	entities requires a license agreement (See http://www.isb-sib.ch/announcement/
RRP	or send an email to license@isb-sib.ch).
RRP	-----
RRP	EMBL: D10077; BAA00971.1; -
DR	InterPro: IPR002522; HCV_capsid.
DR	InterPro: IPR002521; HCV_core.
DR	InterPro: IPR002519; HCV_env.
DR	InterPro: IPR002531; HCV_NS1.
DR	Pfam: PF01543; HCV_capsid; 1.

DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSMEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;
 Query Match 84.1%; Score 127; DB 1; Length 737;
 Best Local Similarity 92.3%; Pred. No. 5e-11;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPORKTKRNTIRRPQDVKKFPGGVI 26
 DB 5 PKPORKTKRNTIRRPQDVKKFPGGQI 30
 RESULT 7
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5C (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits

phosphorylation mediated by cAMP-dependent protein kinase.";
 Eur. J. Biochem. 237:611-618(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moonaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 Cell 87:331-342(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of HX strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M58335; AAA72945.1; -
 DR PIR; A38465; GNVVTC.
 DR PDB; 1A10; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QUV; 26-JUN-00.
 DR PDB; 80HW; 20-APR-99.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4A.
 DR InterPro; IPR001490; HCV_NS4B.
 DR InterPro; IPR002868; HCV_NS5A.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.

DR PFam; PF01001; HCV_NS4b; 1.
 DR PFam; PF01506; HCV_NS5a; 1.
 DR PFam; PF00998; Viral_RDRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1039 1047 STRAND
 FT STRAND 1050 1050 STRAND
 FT STRAND 1059 1063 STRAND
 FT STRAND 1068 1074 STRAND
 FT TURN 1075 1076 TURN
 FT STRAND 1077 1081 STRAND
 FT HELIX 1082 1085 HELIX
 FT TURN 1086 1087 TURN
 FT STRAND 1090 1092 STRAND
 FT TURN 1093 1094 TURN
 FT STRAND 1095 1097 STRAND
 FT STRAND 1101 1103 STRAND
 FT TURN 1104 1107 TURN
 FT STRAND 1108 1112 STRAND
 FT STRAND 1120 1120 STRAND
 FT STRAND 1122 1122 STRAND
 FT STRAND 1129 1133 STRAND
 FT TURN 1135 1136 TURN
 FT STRAND 1139 1144 STRAND
 FT STRAND 1149 1157 STRAND
 FT HELIX 1158 1161 HELIX
 FT TURN 1162 1163 TURN
 FT TURN 1165 1166 TURN
 FT STRAND 1168 1171 STRAND
 FT TURN 1172 1174 TURN
 FT STRAND 1175 1186 STRAND

FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFFD9C CRC64;
 Query Match 84.1%; Score 127; DB 1; Length 3010;
 Best Local Similarity 92.3%; Pred. No. 2.3e-10;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQRTKNTTIRPDQVKFGGVI 26
 DB 5 PKPQRTKNTTIRPDQVKFGGQI 30
 RESULT 8
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90208; BAA14233.1; --
 DR PIR; A39253; GNWVCU.

DR HSP: P26663: 1JXP.
 DR MEROPS: S29.001: -.
 DR MEROPS: U39.001: -.
 DR InterPro: IPR001410: DEAD.
 DR InterPro: IPR002522: HCV capsid.
 DR InterPro: IPR002521: HCV core.
 DR InterPro: IPR002519: HCV env.
 DR InterPro: IPR002531: HCV NS1.
 DR InterPro: IPR002518: HCV NS2.
 DR InterPro: IPR004109: HCV NS3.
 DR InterPro: IPR000745: HCV NS4a.
 DR InterPro: IPR001490: HCV NS4b.
 DR InterPro: IPR002868: HCV NS5a.
 DR InterPro: IPR002166: HCV NS5b.
 DR InterPro: IPR001650: Helicase_C.
 DR InterPro: IPR007095: RNA_pol_DS_PS.
 DR InterPro: IPR007094: RNA_pol_PsVr.
 DR pfam: PF01543: HCV capsid; 1.
 DR pfam: PF01542: HCV core; 1.
 DR pfam: PF01539: HCV env; 1.
 DR pfam: PF01560: HCV NS1; 1.
 DR pfam: PF01538: HCV NS2; 1.
 DR pfam: PF02907: HCV NS3; 1.
 DR pfam: PF01006: HCV NS4a; 1.
 DR pfam: PF01001: HCV NS4b; 1.
 DR pfam: PF01506: HCV NS5a; 1.
 DR pfam: PF00271: Helicase_C; 1.
 DR pfam: PF00998: Viral_RGRP; 1.
 DR ProDom: PD186062: HCV NS1; 1.
 DR SMART: SM00487: DEXDC; 1.
 KW polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA: 327017 MW: AA993794F46DB185 CRC64;

Query Match 84.1%; Score 127; DB 1: Length 3010;
 Best Local Similarity 92.3%; Pred. No. 2.3e-10;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
 DB 5 PKPQKTKRNTIRRPQDVKFPGGGVI 30
 RESULT 9
 POLG_HCVH STANDARD: PRT: 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 FT "Genomic structure of the human prototype strain H of hepatitis C
 FT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 FT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 FT "Hepatitis C virus NS3 RNA helicase domain with a bound
 FT oligonucleotide: the crystal structure provides insights into the mode
 FT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

-----
EMBL: M67463; AAA5534.1; -
PIR: A36814; GNKVCH.
PDB; 1HEI; 25-NOV-98.
PDB; 1AIV; 16-FEB-99.
PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_Capsid.
DR InterPro; IPR002521; HCV_Core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS3a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FT CHAIN 192 383 CAPSID PROTEIN C.
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FT CHAIN 810 1026 PROTEIN P7.
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.
FT CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5B.
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1224 1226
FT TURN 1232 1233
FT TURN 1236 1238
FT TURN 1239 1246
FT TURN 1247 1248
FT STRAND 1251 1255
FT HELIX 1258 1271
FT TURN 1272 1272
FT STRAND 1277 1280
FT TURN 1281 1282
FT STRAND 1283 1285
FT STRAND 1291 1295
FT HELIX 1296 1301
FT TURN 1302 1303
FT STRAND 1312 1316
FT TURN 1317 1319
FT HELIX 1323 1335
FT TURN 1336 1340
FT STRAND 1343 1347
FT TURN 1352 1353
FT TURN 1360 1361
FT STRAND 1362 1366
FT STRAND 1368 1368
FT STRAND 1373 1375
FT TURN 1376 1377
FT STRAND 1378 1380
FT HELIX 1382 1385
FT STRAND 1389 1393
FT HELIX 1397 1409
FT TURN 1410 1411
FT STRAND 1414 1417
FT TURN 1419 1420
FT STRAND 1432 1436
FT TURN 1438 1439
FT STRAND 1450 1453
FT STRAND 1456 1463
FT STRAND 1471 1478
FT STRAND 1480 1480
FT HELIX 1481 1488
FT TURN 1489 1490
FT STRAND 1497 1501
FT STRAND 1507 1507
FT STRAND 1511 1511
FT HELIX 1514 1527
FT HELIX 1532 1544
FT STRAND 1550 1550
FT HELIX 1555 1564
FT HELIX 1570 1578
FT TURN 1579 1580
FT HELIX 1584 1597
FT TURN 1598 1598
FT HELIX 1606 1611
FT TURN 1614 1618
FT STRAND 1622 1623
FT STRAND 1627 1627
FT STRAND 1635 1636
FT HELIX 1640 1652
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 84.1%; Score 127; DB 1; Length 3011;
Best Local Similarity 92.3%; Pred. No. 2.3e-10;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRKTKRNTNRRPDVRFPGGGVI 26
Db 5 PKPQRKTKRNTNRRPDVRFPGGGQI 30
|||||
|||||

RESULT 10
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
ID POLG_HCVJ6

```

P26660:
 01-AUG-1992 (Rel. 23, Last sequence update)
 01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 Hepatitis C virus (isolate HC-J6) (HCV).
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 NCBI_TaxID-11113;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE-92044440; PubMed-1658196;
 Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.;
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 J. Gen. Virol. 72:2697-2704(1991).
 -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 (RNA)(N).
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND RNA.
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: D00944; BAA00792.1; -
 PIR: JQ1303; JQ1303.
 HSP: P27958; LHEI.
 MEROPS: S29.001; -
 MEROPS: U39.001; -
 InterPro: IPR001410; DEAD.
 InterPro: IPR002522; HCV_capsid.
 InterPro: IPR002521; HCV_core.
 InterPro: IPR002519; HCV_env.
 InterPro: IPR002531; HCV_NS1.
 InterPro: IPR002518; HCV_NS2.
 InterPro: IPR004109; HCV_NS3.
 InterPro: IPR000745; HCV_NS4a.
 InterPro: IPR001490; HCV_NS4b.
 InterPro: IPR002868; HCV_NS5a.
 InterPro: IPR002166; HCV_RdRP.
 InterPro: IPR001650; Helicase_C.
 InterPro: IPR007095; RNA_pol_DS_PS.
 InterPro: IPR007094; RNA_pol_PSVir.
 Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01535; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.

DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural
 removed from CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT CHAIN 3033 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 84.1%; Score 127; DB 1; Length 3033;
 Best Local Similarity 92.3%; Pred. No. 2.3e-10;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 PKPQKTKRNTIRPQDVKFPGGVI 26
 Db 5 PKPQKTKRNTIRPQDVKFPGGQI 30
 |||||
 1 PKPQKTKRNTIRPQDVKFPGGVI 26
 5 PKPQKTKRNTIRPQDVKFPGGQI 30
 |||||
 RESULT 11
 POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 ID POLG_HCVJ8
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate HC-J8) (HCV).
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID-11115;
 RN [1]

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M84754; -; NOT_ANNOTATED_CDS.

PIR; A40244; GNMVTM.

PDB; 1N64; 25-FEB-03.

PDB; 1NS3; 08-APR-98.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

InterPro: IPR001410; DEAD.

InterPro: IPR002522; HCV_capsid.

InterPro: IPR002521; HCV_core.

InterPro: IPR002519; HCV_env.

InterPro: IPR002531; HCV_NS1.

InterPro: IPR002518; HCV_NS2.

InterPro: IPR004109; HCV_NS3.

InterPro: IPR000745; HCV_NS4a.

InterPro: IPR001490; HCV_NS4b.

InterPro: IPR002868; HCV_NS5a.

InterPro: IPR002166; HCV_RdRp.

InterPro: IPR007095; RNA_pol_DS_PS.

InterPro: IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRp; 1.

DR ProDom; PD186062; HCV_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

3D-structure. 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

INIT_MET 1 1 CELLULAR AMINOPEPTIDASE.

CHAIN 1 115 CORE PROTEIN (POTENTIAL).

FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 347 369 POTENTIAL.

FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 79.5%; Score 120; DB 1; Length 3010;

Best Local Similarity 92.0%; Pred. No. 2.5e-09;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 KPORKTKRNTIRRPQDKVPPGGGVI 26

DB 6 KPORKTKRNTIRRPQDKVPPGGGQI 30

RESULT 13

POLG_HCV1 STANDARD; PRT; 3011 AA.

ID AC P26684;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate 1) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OC NCBI_Taxid=11104;

[1]

SEQUENCE FROM N.A.

RX MEDLINE-91172826; PubMed-1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,

RA Bradley D.W., Kuo G., Houghton M.;

RT "Genetic organization and diversity of the hepatitis C virus.;"

RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +

CC (RNA)(N).

CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M62321; AAA45676.1; -.


```
DR PIR; A39166; GNVWC3..
DR PDB; 1A1V; 16-FEB-99.
DR PDB; 1HEI; 25-NOV-98.
DR MEROPS; S29, 001; -.
DR MEROPS; U39, 001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RGRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00598; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure. 1
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 393
FT CHAIN 394 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT CHAIN 3012 347
FT TRANSMEM 347 369
FT ACT_SITE 1083 1093
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 393
FT CHAIN 394 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT CHAIN 3012 347
FT TRANSMEM 347 369
FT ACT_SITE 1083 1093
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
PIR; A39166; GNVWC3..
PDB; 1A1V; 16-FEB-99.
PDB; 1HEI; 25-NOV-98.
MEROPS; S29, 001; -.
MEROPS; U39, 001; -.
InterPro; IPR001410; DEAD.
InterPro; IPR002522; HCV_capsid.
InterPro; IPR002521; HCV_core.
InterPro; IPR002519; HCV_env.
InterPro; IPR002531; HCV_NS1.
InterPro; IPR002518; HCV_NS2.
InterPro; IPR004109; HCV_NS3.
InterPro; IPR000745; HCV_NS4a.
InterPro; IPR001490; HCV_NS4b.
InterPro; IPR002868; HCV_NS5a.
InterPro; IPR002166; HCV_RGRP.
InterPro; IPR001650; Helicase_C.
InterPro; IPR007095; RNA_pol_DS_Ps.
InterPro; IPR007094; RNA_pol_PSVir.
Pfam; PF01543; HCV_capsid; 1.
Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_env; 1.
Pfam; PF01560; HCV_NS1; 1.
Pfam; PF01538; HCV_NS2; 1.
Pfam; PF02907; HCV_NS3; 1.
Pfam; PF01006; HCV_NS4a; 1.
Pfam; PF01001; HCV_NS4b; 1.
Pfam; PF01506; HCV_NS5a; 1.
Pfam; PF00271; Helicase_C; 1.
Pfam; PF00598; Viral_RdRp; 1.
ProDom; PD186062; HCV_NS1; 1.
SMART; SM00487; DEXDC; 1.
Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
3D-structure. 1
INIT_MET 1
CHAIN 1 115
CHAIN 116 191
CHAIN 192 393
CHAIN 394 729
CHAIN 730 1006
CHAIN 1007 1615
CHAIN 1616 1862
CHAIN 1863 2013
CHAIN 2014 3011
CHAIN 3012 347
TRANSMEM 347 369
ACT_SITE 1083 1093
ACT_SITE 1107 1107
ACT_SITE 1165 1165
NP_BIND 1230 1237
SITE 1316 1319
CARBOHYD 196 196
CARBOHYD 209 209
CARBOHYD 234 234
CARBOHYD 305 305
CARBOHYD 417 417
CARBOHYD 423 423
CARBOHYD 430 430
CARBOHYD 448 448
CARBOHYD 476 476
CARBOHYD 532 532
CARBOHYD 540 540
CARBOHYD 556 556
CARBOHYD 576 576
CARBOHYD 623 623
CARBOHYD 645 645
CARBOHYD 2041 2041
CARBOHYD 2077 2077
CARBOHYD 2240 2240
CARBOHYD 2364 2364
CARBOHYD 2789 2789
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
Query Match 78.8%; Score 119; DB 1; Length 3011;
Best Local Similarity 84.6%; Pred. No. 3.5e-09;
Matches 22; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
| | | | | | | | | | | | | | | |
DB 5 PKPQKTKRNTIRRPQDVKFGGGGI 30
| | | | | | | | | | | | | | | |
RESULT 14
ID RK4_TOBAC STANDARD; PRT; 282 AA.
AC O80361;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L4, chloroplast precursor (R-protein L4).
OS RPL4.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bright yellow 4; TISSUE=Leaf;
RA Yokoi F., Ohta M., Sugiyama M.;
RT "Tobacco chloroplast ribosomal protein L4.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY AND SPECIFICALLY TO 23S RRNA
CC (BY SIMILARITY). MAY PLAY A ROLE IN PLASTID TRANSCRIPTIONAL
CC REGULATION.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE L4P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB010878; BAA31510.1; -.
CC PIR; T01739; T01739.
CC InterPro; IPR002136; Ribosomal_L4/L1E.
CC Pfam; PF00573; Ribosomal_L4; 1.
KW Ribosomal protein; rRNA-binding; Chloroplast; Transit peptide.
FT TRANSIT 1 43 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 44 282 50S RIBOSOMAL PROTEIN L4.
FT DOMAIN 254 282 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 282 AA; 31252 MW; 12BFAC9C08D9E4AC CRC64;
Query Match 35.1%; Score 53; DB 1; Length 282;
Best Local Similarity 43.3%; Pred. No. 1.7;
Matches 13; Conservative 5; Mismatches 6; Indels 6; Gaps 2;
QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 27
| | | | | | | | | | | | | | | |
DB 108 PYPQKTKRARRGSRNRTPLR---PGGGVVF 134
| | | | | | | | | | | | | | | |
RESULT 15
ID FURI_HUMAN STANDARD; PRT; 794 AA.
AC P09958; Q14336;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Furin precursor (EC 3.4.21.75) (Paired basic amino acid residue
DE cleaving enzyme) (PACE) (Dibasic processing enzyme).
```

GN FURIN OR PACE OR FUR.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=90175002; PubMed=2408021;
 RA van den Ouweland A.M.W., van Duijnhoven H.L.P., Kelzer G.D.,
 RA Dorssers L.C.J., van de Ven W.J.M.;
 RT *Structural homology between the human fur gene product and the
 RT subtilisin-like protease encoded by yeast KEX2.*;
 RL Nucleic Acids Res. 18:664-664(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91321735; PubMed=1713771;
 RA Barr P.J., Mason O.B., Landsberg K.E., Wong P.A., Kiefer M.C.,
 RA Brake A.J.;
 RT *DNA and gene structure for a human subtilisin-like protease with
 RT cleavage specificity for paired basic amino acid residues.*;
 RL DNA Cell Biol. 10:319-328(1991).
 RN [3]
 RP SEQUENCE OF 296-794 FROM N.A.
 RX MEDLINE=87053858; PubMed=3023061;
 RA Roebroek A.J.M., Schaiken J.A., Leunissen J.A.M., Onnekink C.,
 RA Bloemers H.P.J., van de Ven W.J.M.;
 RT *Evolutionary conserved close linkage of the c-fes/fps proto-oncogene
 RT and genetic sequences encoding a receptor-like protein.*;
 RL EMBO J. 5:2197-2202(1986).
 RN [4]
 RP 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.
 RX MEDLINE=94291619; PubMed=8020465;
 RA Slezan R.J., Creemers J.W.M., van de Ven W.J.M.;
 RT *Homology modelling of the catalytic domain of human furin. A model
 RT for the eukaryotic subtilisin-like proprotein convertases.*;
 RL Eur. J. Biochem. 222:255-266(1994).
 RN [5]
 RP PROCESSING.
 RX MEDLINE=92332543; PubMed=1629222;
 RA Leduc R., Molloy S.S., Thorne B.A.;
 RT *Activation of human furin precursor processing endoprotease occurs by
 RT an intramolecular autoproteolytic cleavage.*;
 RL J. Biol. Chem. 267:14304-14308(1992).
 CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTIDASE
 CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
 CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-1-2aa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
 CC complement component C3 and von Willebrand factor from their
 CC respective precursors.
 CC -1- COFACTOR: CALCIUM-DEPENDENT.
 CC -1- ENZYME REGULATION: COULD BE INHIBITED BY THE NOT SECONDLY CLEAVED
 CC PROPEPTIDE.
 CC -1- SUBCELLULAR LOCATION: SEEMS TO BE LOCALIZED INTRACELLULARLY TO THE
 CC TRANS GOLGI NETWORK. PROPEPTIDE CLEAVAGE IS A PREREQUISITE FOR
 CC EXIT OF FURIN MOLECULES OUT OF THE ENDOPLASMIC RETICULUM (ER).
 CC SECOND CLEAVAGE IN THE PROPEPTIDE OCCUR IN THE TRANS GOLGI NETWORK
 CC (TGN), WHICH IS FOLLOWED BY THE RELEASE OF THE PROPEPTIDE BOUND TO
 CC FURIN AND THE ACTIVATION OF FURIN.
 CC -1- TISSUE SPECIFICITY: SEEMS TO BE EXPRESSED UBIQUITOUSLY.
 CC -1- DOMAIN: CONTAINS A CYTOPLASMIC DOMAIN RESPONSIBLE FOR ITS TGN
 CC LOCALIZATION AND RECYCLING FROM THE CELL SURFACE.
 CC -1- PTM: THE PROPEPTIDE IS AUTOCATALYTICALLY REMOVED THROUGH AN
 CC INTRAMOLECULAR CLEAVAGE PROBABLY IN THE ENDOPLASMIC RETICULUM
 CC (ER). IN THE TGN THE SECOND CLEAVAGE IN THE PROPEPTIDE COULD LEAD
 CC TO THE ACTIVATION OF FURIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC -1- SIMILARITY: Contains 1 homo B/P domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X17094; CAA34948.1; -;
 DR EMBL; X04329; CAA27860.1; -;
 DR EMBL; A06939; CAA00605.1; -;
 DR PIR; A39552; KXHUF.
 DR HSP; Q99405; LMPT.
 DR MEROPS; S08.071; -;
 DR Genew; HGNC:8566; FURIN.
 DR MIM; 136950; -;
 DR GO; GO:0005794; C:Golgi apparatus; TAS.
 DR GO; GO:0004276; F:furin activity; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
 DR InterPro; IPR006212; Furin_repeat.
 DR InterPro; IPR002884; P_domain.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF01483; P_protein; PARTIAL.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00261; FU; 2.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Glycoprotein; Signal;
 KW Zymogen; Calcium.
 FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 107
 FT CHAIN 108 794 FURIN.
 FT DOMAIN 556 705
 FT TRANSEM 716 738
 FT ACT_SITE 153 194
 FT ACT_SITE 194 194
 FT ACT_SITE 368 368
 FT ACT_SITE 211 360
 FT DISULFID 303 333
 FT CARBOHYD 387 387
 FT CARBOHYD 440 440
 FT CARBOHYD 553 553
 FT SITE 70 75
 FT SITE 104 107
 FT SITE 498 500
 FT SITE 759 762
 FT SITE 773 779
 SQ SEQUENCE 794 AA; 86678 MW; 10C44DD5892EF85D CRC64;
 Query Match 34.4%; Score 52; DB 1; Length 794;
 Best Local Similarity 50.0%; Pred. No. 7.3;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 4 QRKTKRNTIRRPQDVKFP 21
 Db 102 KRTRKRVQVQETDPKFP 119
 Search completed: August 7, 2003, 11:20:03
 Job time : 5.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-30

Perfect score: 151
Sequence: 1 PKPQKTKRNTIRRPQDVKEFGGGVIYV 28

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues

total number of hits	satisfying chosen parameters:	830525
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

8
7
4
5
1
3
2
2
1
1
0
0
1
0
4
5
0

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23: *

```

1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	151	100.0	3021	12	Q81495	Q81495	hepatitis c
2	134	88.7	45	12	Q68307	Q68307	hepatitis c
3	134	88.7	45	12	Q68308	Q68308	hepatitis c
4	134	88.7	45	12	Q68310	Q68310	hepatitis c
5	134	88.7	60	12	Q8JRS2	Q8JRS2	hepatitis c
6	134	88.7	61	12	Q8JYS3	Q8JYS3	hepatitis c
7	134	88.7	73	12	Q8JYR4	Q8JYR4	hepatitis c
8	134	88.7	100	12	Q8QP85	Q8QP85	hepatitis c
9	134	88.7	100	12	Q8QP87	Q8QP87	hepatitis c
10	134	88.7	100	12	Q8QP72	Q8QP72	hepatitis c
11	134	88.7	100	12	Q8QP71	Q8QP71	hepatitis c
12	134	88.7	100	12	Q8QP74	Q8QP74	hepatitis c
13	134	88.7	100	12	Q8QP80	Q8QP80	hepatitis c
14	134	88.7	100	12	Q8QP77	Q8QP77	hepatitis c
15	134	88.7	100	12	Q8QP84	Q8QP84	hepatitis c
16	134	88.7	100	12	Q8QP75	Q8QP75	hepatitis c

17	134	88.7	100	12	Q8p79	hepatitis c
18	134	88.7	100	12	Q8p81	hepatitis c
19	134	88.7	100	12	Q8p83	hepatitis c
20	134	88.7	100	12	Q8p76	hepatitis c
21	134	88.7	100	12	Q8p76	hepatitis c
22	134	88.7	109	12	Q8p86	hepatitis c
23	134	88.7	109	12	Q81340	hepatitis c
24	134	88.7	109	12	Q81341	hepatitis c
25	134	88.7	114	12	Q8m374	hepatitis c
26	134	88.7	114	12	Q68892	hepatitis c
27	134	88.7	114	12	Q68893	hepatitis c
28	134	88.7	119	12	Q8BCX2	hepatitis c
29	134	88.7	119	12	Q8BCX1	hepatitis c
30	134	88.7	119	12	Q8BCX0	hepatitis c
31	134	88.7	119	12	Q8BCW9	hepatitis c
32	134	88.7	119	12	Q8BCW8	hepatitis c
33	134	88.7	119	12	Q8BCW7	hepatitis c
34	134	88.7	119	12	Q8BCW6	hepatitis c
35	134	88.7	119	12	Q8BCW4	hepatitis c
36	134	88.7	122	12	Q8QRJ1	hepatitis c
37	134	88.7	122	12	Q8QM16	hepatitis c
38	134	88.7	122	12	Q8QM18	hepatitis c
39	134	88.7	122	12	Q8QRJ4	hepatitis c
40	134	88.7	122	12	Q8QRJ5	hepatitis c
41	134	88.7	122	12	Q8QM15	hepatitis c
42	134	88.7	122	12	Q8QM19	hepatitis c
43	134	88.7	122	12	Q8QM17	hepatitis c
44	134	88.7	122	12	Q8QM10	hepatitis c
45	134	88.7	122	12	Q8QRJ3	hepatitis c

ALIGNMENTS

RESULT 1

```

AC00491 1
ID AC Q81495 PRELIMINARY; PRT; 3021 AA.
DC AC Q81495;
DT DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DE Genome polyprotein.
OS OS Hepatitis C virus.
OC OC Viruses; ssRNA positive-strand viruses, no DNA stage; FL
OC OC Hepacivirus.
OX OX NCBI_Taxid=11103;
RN RN [1]
RT RP SEQUENCE FROM N.A.
RC RC STRAIN-K3a;
RA RA Date T.;
RL RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN RN [2]
RT RP SEQUENCE FROM N.A.
RC RC STRAIN-K3a;
RA RA Yamada N., Manihara K., Mizokami M., Ohba K., Takada A.,
RA RA Date T.;
RL RL "Full-length sequence of the genome of hepatitis C virus
RT RT comparative study with different genotypes.";
RN RN J. Gen. Virol. 75:3279-3284(1994);
CC CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID
CC CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO P
CC CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A P
CC CC PROTEIN C AND MNA (BY SIMILARITY).
RN RN EMBL; D28917; BAA06044.1; -
DR DR HSP; P27958; 1HEI.
DR DR MEROPS; S29.001; -.
DR DR MEROPS; U39.001; -.
DR DR InterPro; IPR001410; DEAD.
DR DR InterPro; IPR002522; HCV_capsid.
DR DR InterPro; IPR002521; HCV_core.
DR DR InterPro; IPR002519; HCV_env.
DR DR InterPro; IPR002531; HCV_NSI.

```

```

DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; HCV_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05051; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3021 AA; 328386 MW; A97418FF36C062A4 CRC64;

Query Match 100.0%; Score 151; DB 12; Length 3021;
Best Local Similarity 100.0%; Pred. No. 7.4e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGPGGVIYV 28
DB 5 PKPQRTKTRNTIRRPQDVKFGPGGVIYV 32

RESULT 2
Q68307 ID Q68307 PRELIMINARY; PRT; 45 AA.
AC Q68307;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
STRAIN-HCV-BB21;
Songvilalai S., Kanistanon D., Kunkitti R.;
"Identification and characterisation of Thai isolates of hepatitis C
virus.";
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: U23745; AAA65052.1;
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
Polyprotein.
NON_TER 45 45
SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 88.7%; Score 134; DB 12; Length 45;
Best Local Similarity 96.2%; Pred. No. 3.3e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGPGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKFGPGGVI 30

RESULT 4
Q68310 ID Q68310 PRELIMINARY; PRT; 45 AA.
AC Q68310;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
STRAIN-HCV-C94009;
Songvilalai S., Kanistanon D., Kunkitti R.;
"Identification and characterisation of Thai isolates of hepatitis C
virus.";
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: U23748; AAA65055.1;
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
Polyprotein.
NON_TER 45 45
SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 88.7%; Score 134; DB 12; Length 45;
Best Local Similarity 96.2%; Pred. No. 3.3e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGPGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKFGPGGVI 30

```

```

RESULT 3
Q68308 ID Q68308 PRELIMINARY; PRT; 45 AA.
AC Q68308;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
STRAIN-HCV-BB38;
Songvilalai S., Kanistanon D., Kunkitti R.;
"Identification and characterisation of Thai isolates of hepatitis C
virus.";
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: U23746; AAA65053.1;
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
Polyprotein.
NON_TER 45 45
SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 88.7%; Score 134; DB 12; Length 45;
Best Local Similarity 96.2%; Pred. No. 3.3e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGPGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKFGPGGVI 30

RESULT 4
Q68310 ID Q68310 PRELIMINARY; PRT; 45 AA.
AC Q68310;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
STRAIN-HCV-C94009;
Songvilalai S., Kanistanon D., Kunkitti R.;
"Identification and characterisation of Thai isolates of hepatitis C
virus.";
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: U23748; AAA65055.1;
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
Polyprotein.
NON_TER 45 45
SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 88.7%; Score 134; DB 12; Length 45;
Best Local Similarity 96.2%; Pred. No. 3.3e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKFGPGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKFGPGGVI 30

```

```

QY 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKPPGGGQI 30

RESULT 5
Q8JYS2 PRELIMINARY; PRT; 60 AA.
AC Q8JYS2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG253;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF506613; AAM33389.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein. 60
KW NON_TER
FT SEQUENCE 60 AA; 6725 MW; ACCE7D9C8B90299A CRC64;

Query Match 88.7%; Score 134; DB 12; Length 60;
Best Local Similarity 96.2%; Pred. No. 4.5e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKPPGGGQI 30

RESULT 6
Q8JYS3 PRELIMINARY; PRT; 61 AA.
AC Q8JYS3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG256;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF506612; AAM33388.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
KW CHAIN
FT NON_TER 1 >61 CORE PROTEIN.
FT SEQUENCE 61 AA; 6881 MW; B92CCE7D9C8B9029 CRC64;

```

```

Query Match 88.7%; Score 134; DB 12; Length 61;
Best Local Similarity 96.2%; Pred. No. 4.6e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
DB 5 PKPQRTKTRNTIRRPQDVKPPGGGQI 30

RESULT 7
Q8JYR4 PRELIMINARY; PRT; 73 AA.
AC Q8JYR4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG291;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF506621; AAM33397.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
KW NON_TER 1 1
FT NON_TER 73 73
FT SEQUENCE 73 AA; 8201 MW; 6EDC082DAE8CB796 CRC64;

Query Match 88.7%; Score 134; DB 12; Length 73;
Best Local Similarity 96.2%; Pred. No. 5.6e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
DB 1 PKPQRTKTRNTIRRPQDVKPPGGGQI 26

RESULT 8
Q8QP85 PRELIMINARY; PRT; 100 AA.
AC Q8QP85;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-602;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).

```

```
DR EMBL: AY070180; AAL58594.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polypeptide.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQI 30

RESULT 9
Q8QP72 PRELIMINARY; PRT; 100 AA.
Q8QP72:
01-JUN-2002 (TRENBLrel. 21, Created)
01-JUN-2002 (TRENBLrel. 21, Last sequence update)
01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=233;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg."
RL J. Virol. 76:4034-4043(2002).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AY070178; AAL58592.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polypeptide.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQI 30

RESULT 10
Q8QP72 PRELIMINARY; PRT; 100 AA.
Q8QP72:
01-JUN-2002 (TRENBLrel. 21, Created)
01-JUN-2002 (TRENBLrel. 21, Last sequence update)
01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=609;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
```

```
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg."
RL J. Virol. 76:4034-4043(2002).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AY070193; AAL58607.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polypeptide.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQI 30

RESULT 11
Q8QP71 PRELIMINARY; PRT; 100 AA.
Q8QP71:
01-JUN-2002 (TRENBLrel. 21, Created)
01-JUN-2002 (TRENBLrel. 21, Last sequence update)
01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=637;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg."
RL J. Virol. 76:4034-4043(2002).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AY070194; AAL58608.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polypeptide.
FT NON_TER
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFFGGGQI 30

RESULT 12
Q8QP74 PRELIMINARY; PRT; 100 AA.
Q8QP74:
01-JUN-2002 (TRENBLrel. 21, Created)
01-JUN-2002 (TRENBLrel. 21, Last sequence update)
01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
```

```

OC Hepacivirus
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=631;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070191; AAL58605.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
SQ SEQUENCE 100 AA; 11289 MW; F3BC90F5F595F18B CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFGGGQI 30

RESULT 13
QSQP80
ID Q8QP80 PRELIMINARY; PRT; 100 AA.
AC Q8QP80;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=781;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070185; AAL58599.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
SQ SEQUENCE 100 AA; 11289 MW; F3BC90F5F595F18B CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFGGGQI 30

RESULT 14
QSQP77
ID Q8QP77 PRELIMINARY; PRT; 100 AA.
AC Q8QP77;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=718;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070181; AAL58595.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F585AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFGGGQI 30

RESULT 15
QSQP84
ID Q8QP84 PRELIMINARY; PRT; 100 AA.
AC Q8QP84;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=718;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070181; AAL58595.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F585AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFGGGQI 30

```

```

AC Q8QP77;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=748;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070188; AAL58602.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
SQ SEQUENCE 100 AA; 11289 MW; F3BC90F5F595F18B CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFGGGQI 30

RESULT 15
QSQP84
ID Q8QP84 PRELIMINARY; PRT; 100 AA.
AC Q8QP84;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=718;
RX MEDLINE=21904745; PubMed=11907242;
RA Kalinina O., Norder H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070181; AAL58595.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 100
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F585AE6 CRC64;

Query Match      88.7%; Score 134; DB 12; Length 100;
Best Local Similarity 96.2%; Pred. No. 7.8e-13;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFGGGVI 26
DB 5 PKPQKTKRNTIRRPQDVKFGGGQI 30

```

Db 5 PKPQRTKRNTIRRPQDVKEPGGQI 30
|||||

Search completed: August 7, 2003, 11:19:02
Job time : 26.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-30

Perfect score: 151

Sequence: 1 PKPQKTKRNTIRPDQKFGGGVIIV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

1 number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/laa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/laa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/laa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/laa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/laa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/laa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	100.0	28	3	US-08-921-887-30
2	134	88.7	28	3	US-08-921-887-23
3	134	88.7	191	2	US-08-290-665A-187
4	134	88.7	191	2	US-08-290-665A-188
5	134	88.7	191	2	US-08-290-665A-190
6	134	88.7	191	5	PCT-US95-10398-187
7	134	88.7	191	5	PCT-US95-10398-188
8	134	88.7	191	5	PCT-US95-10398-190
9	134	88.7	319	3	US-08-836-075A-44
10	134	88.7	319	4	US-08-635-886C-230
11	133	88.1	191	2	US-08-290-665A-189
12	133	88.1	191	5	PCT-US95-10398-189
13	129	85.4	28	3	US-08-921-887-25
14	129	85.4	450	4	US-08-635-886C-191
15	129	85.4	450	4	US-08-635-886C-192
16	128	84.8	191	2	US-08-290-665A-175
17	128	84.8	191	5	PCT-US95-10398-175
18	127	84.1	30	1	US-08-324-977-6
19	127	84.1	30	2	US-08-384-616-6
20	127	84.1	30	2	US-08-904-686A-6
21	127	84.1	30	3	US-09-315-850-6
22	127	84.1	31	1	US-07-681-701-8
23	127	84.1	34	3	US-08-380-160-6
24	127	84.1	43	4	US-09-020-846-36
25	127	84.1	44	3	US-08-380-160-2
26	127	84.1	44	4	US-09-389-756-1
27	127	84.1	45	3	US-08-380-160-1

28 127 84.1 61 1 US-07-945-054-9 Sequence 9, Appli
29 127 84.1 61 1 US-08-083-947-23 Sequence 23, Appli
30 127 84.1 61 1 US-08-530-530-3 Sequence 3, Appli
31 127 84.1 61 1 US-08-262-037-26 Sequence 26, Appli
32 127 84.1 61 5 PCT-US94-08638-9 Sequence 9, Appli
33 127 84.1 61 5 PCT-US94-07088-23 Sequence 23, Appli
34 127 84.1 61 5 PCT-US95-13660-3 Sequence 3, Appli
35 127 84.1 74 3 US-08-836-075A-10 Sequence 10, Appli
36 127 84.1 74 4 US-08-635-886C-198 Sequence 198, App
37 127 84.1 100 4 US-08-635-886C-232 Sequence 232, App
38 127 84.1 108 3 US-08-836-075A-14 Sequence 14, Appli
39 127 84.1 115 1 US-08-324-977-8 Sequence 8, Appli
40 127 84.1 115 2 US-08-384-616-8 Sequence 8, Appli
41 127 84.1 115 2 US-08-904-686A-8 Sequence 8, Appli
42 127 84.1 115 3 US-09-315-850-8 Sequence 2, Appli
43 127 84.1 123 2 US-08-501-195-2 Sequence 15, Appli
44 127 84.1 124 1 US-08-244-116B-15 Sequence 46, Appli
45 127 84.1 137 3 US-08-836-075A-46

ALIGNMENTS

RESULT 1

US-08-921-887-30
; Sequence 30, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-30

Query Match 100.0%; Score 151; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.5e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGVIV 28
 DB 1 PKPQKTKRNTIRRPQDVKFPGGVIV 28

RESULT 2

US-08-921-887-23
 ; Sequence 23, Application US/08921887
 ; Patent No. 6030771
 ; GENERAL INFORMATION:
 ; APPLICANT: KHUYAKOV, YURI E.
 ; APPLICANT: FIELDS, HOWARD A.
 ; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
 ; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
 ; NUMBER OF SEQUENCES: 55
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: JONES & ASKEW, LLP
 ; STREET: 191 Peachtree Street, N.W., 37th Floor
 ; CITY: Atlanta
 ; STATE: GA
 ; COUNTRY: USA
 ; ZIP: 30303-1769
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/921.887
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WARREN, WILLIAM L.
 ; REGISTRATION NUMBER: 36,714
 ; REFERENCE/DOCKET NUMBER: 03063-0380
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 404-818-3700
 ; TELEFAX: 404-818-3799
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 28 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Hepatitis virus
 ; 18-921-887-23

Query Match 88.7%; Score 134; DB 3; Length 28;
 Best Local Similarity 96.2%; Pred. No. 4.5e-13;
 Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGVIV 26
 DB 1 PKPQKTKRNTIRRPQDVKFPGGVIV 26

RESULT 3

US-08-290-665A-187
 ; Sequence 187, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R.H. AND
 ; APPLICANT: PURCELL, R.H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 187:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 191 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; ORIGINAL SOURCE:
 ; ORGANISM: homomapsiens
 ; INDIVIDUAL ISOLATE: HK10
 ; US-08-290-665A-187

Query Match 88.7%; Score 134; DB 2; Length 191;
 Best Local Similarity 96.2%; Pred. No. 3.6e-12;
 Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGVIV 26
 DB 5 PKPQKTKRNTIRRPQDVKFPGGVIV 30

RESULT 4

US-08-290-665A-188
 ; Sequence 188, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R.H. AND
 ; APPLICANT: PURCELL, R.H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994

Qy 1 PKPQRKTKRNTIRRPQDVKFPGGGVI 26
|||
Db 5 PKPQRKTKRNTIRRPQDVKFPGGGQI 30
|||

```

RESULT 7
PCT-US95-10398-188
; Sequence 188, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S52
PCT-US95-10398-188

Query Match 88.7%; Score 134; DB 5; Length 191;
Best Local Similarity 96.2%; Pred. No. 3.6e-12;
Matches 25; Conservative 0; Mismatches 1; Indels 0;

QY 1 PKPORKTKRNTIRRPDVKFPGGGVI 26
| | | | | | | | | | | | | | | | | | | | |
DB 5 PKPORKTKRNTIRRPDVKFPGGGQI 30

RESULT 8
PCT-US95-10398-190
; Sequence 190, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 319 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-075A-44

Query Match 88.7%; Score 134; DB 3; Length 319;
Best Local Similarity 96.2%; Pred. No. 6.4e-12;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTNIRPDQVKFPGGGVI 26
Db 5 PKPQRTKRTNIRPDQVKFPGGGQI 30

RESULT 10

US-08-635-886C-230
; Sequence 230, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; PRIOR FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 230
; LENGTH: 319
; TYPE: PRT
; ORGANISM: hepatitis C virus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (144)..(144)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (156)..(157)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE

LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (167)..(167)
OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (171)..(172)
OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (174)..(174)
OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (177)..(177)
OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (232)..(232)
OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (233)..(233)
OTHER INFORMATION: Xaa is any amino acid
US-08-635-886C-230

Query Match 88.7%; Score 134; DB 4; Length 319;
Best Local Similarity 96.2%; Pred. No. 6.4e-12;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTNIRPDQVKFPGGGVI 26
Db 5 PKPQRTKRTNIRPDQVKFPGGGQI 30

RESULT 11

US-08-290-665A-189
; Sequence 189, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R. H. AND
; APPLICANT: PURCELL, R. H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849

TELEX: 421792
INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-189

Query Match 88.1%; Score 133; DB 2; Length 191;
Best Local Similarity 92.3%; Pred. No. 5.1e-12;
Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
DB 5 PKPQKTKRNTIRRPQDIKFPGGGQI 30

JUT 12

PCT-US95-10398-189
Sequence 189, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
PCT-US95-10398-189

Query Match 88.1%; Score 133; DB 5; Length 191;
Best Local Similarity 92.3%; Pred. No. 5.1e-12;
Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
DB 5 PKPQKTKRNTIRRPQDIKFPGGGQI 30

RESULT 13

US-08-921-887-25
Sequence 25, Application US/08921887
Patent No. 6030771
GENERAL INFORMATION:
APPLICANT: KHUYAKOV, YURI E.
APPLICANT: FIELDS, HOWARD A.
TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAM
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: JONES & ASKEW, LLP
STREET: 191 Peachtree Street, N.W., 37th Floor
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30303-1769

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/921,887
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WARREN, WILLIAM L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 03063-0380
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis virus

US-08-921-887-25
Query Match 85.4%; Score 129; DB 3; Length 28;
Best Local Similarity 92.3%; Pred. No. 2.4e-12;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTIRRPQDVKFPGGGVI 26
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
DB 1 PKPQKTKRNTYRRPQDVKFPGGGQI 26

RESULT 14

US-08-635-886C-191
Sequence 191, Application US/08635886C
Patent No. 6555114
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert

APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2752-18
CURRENT APPLICATION NUMBER: US/08/635,886C
PRIOR FILING DATE: 1996-04-25
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 191
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-635-886C-191

Query Match 85.4%; Score 129; DB 4; Length 450;

Best Local Similarity 92.3%; Pred. No. 5e-11; 2; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTIRRPQDVKEPFGGVI 26
||||||| |||||||||
DB 5 PKPQRTKRNTIRRPQDVKEPFGGQI 30

RESULT 15
US-08-635-886C-192
Sequence 192, Application US/08635886C
Patent No. 6555114
GENERAL INFORMATION:
APPLICANT: LEROUX-ROELS, Geert
APPLICANT: DELEYS, Robert
APPLICANT: MAERTENS, Geert
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
FILE REFERENCE: 2752-18
CURRENT APPLICATION NUMBER: US/08/635,886C
CURRENT FILING DATE: 1996-04-25
PRIOR APPLICATION NUMBER: PCT/EP94/03555
PRIOR FILING DATE: 1994-10-28
PRIOR APPLICATION NUMBER: EP 93402718.6
PRIOR FILING DATE: 1993-11-04
NUMBER OF SEQ ID NOS: 286
SOFTWARE: PatentIn version 3.1
SEQ ID NO 192
LENGTH: 450
TYPE: PRT
ORGANISM: hepatitis C virus
US-08-635-886C-192

Query Match 85.4%; Score 129; DB 4; Length 450;

Best Local Similarity 92.3%; Pred. No. 5e-11; 2; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTIRRPQDVKEPFGGVI 26
||||||| |||||||||
DB 5 PKPQRTKRNTIRRPQDVKEPFGGQI 30

Search completed: August 7, 2003, 11:23:52
Job time : 10.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146A-30

Perfect score: 151

Sequence: 1 PKPQRTKRTIRPQDVKPPGGVIV 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

tal number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	88.7	319	10	US-09-851-138-44
2	127	84.1	44	15	US-10-367-677-1
3	127	84.1	74	10	US-09-851-138-10
4	127	84.1	91	9	US-09-758-308-1
5	127	84.1	97	10	US-09-756-875-8
6	127	84.1	103	10	US-09-921-397-77
7	127	84.1	108	10	US-09-851-138-14
8	127	84.1	113	10	US-09-921-397-78
9	127	84.1	137	10	US-09-851-138-46
10	127	84.1	138	10	US-09-851-138-60
11	127	84.1	166	11	US-09-899-046-152
12	127	84.1	166	11	US-09-878-281-152
13	127	84.1	169	11	US-09-899-046-42
14	127	84.1	169	11	US-09-899-046-44
15	127	84.1	169	11	US-09-878-281-42
					Sequence 44, Appl
					Sequence 1, Appl
					Sequence 10, Appl
					Sequence 1, Appl
					Sequence 77, Appl
					Sequence 8, Appl
					Sequence 14, Appl
					Sequence 78, Appl
					Sequence 46, Appl
					Sequence 60, Appl
					Sequence 152, App
					Sequence 152, App
					Sequence 42, Appl
					Sequence 44, Appl
					Sequence 42, Appl

Sequence 44, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 76, Appl
Sequence 12, Appl
Sequence 18, Appl
Sequence 50, Appl
Sequence 52, Appl
Sequence 54, Appl
Sequence 144, App
Sequence 50, Appl
Sequence 52, Appl
Sequence 54, Appl
Sequence 144, App
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 40, Appl
Sequence 4, Appl
Sequence 9, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 2, Appl

Sequence 44, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 76, Appl
Sequence 12, Appl
Sequence 18, Appl
Sequence 50, Appl
Sequence 52, Appl
Sequence 54, Appl
Sequence 144, App
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 40, Appl
Sequence 4, Appl
Sequence 9, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-851-138-44
; Sequence 44, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN

AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

CITY: HOUSTON

STATE: TEXAS

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,138

FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>

APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994

APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:
LENGTH: 319 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-851-138-44

Query Match 88.7%; Score 134; DB 10; Length 319;
Best Local Similarity 96.2%; Pred. No. 1.9e-11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTIRRPQDVKPPGGVI 26
|||||
DB 5 PKPQRTKNTIRRPQDVKPPGGQI 30

RESULT 2

US-10-367-677-1
Sequence 1, Application US/10367677
Publication No. US20030118604A1
GENERAL INFORMATION:
APPLICANT: JOLIVET, MICHEL
APPLICANT: PENIN, FRANCOIS
APPLICANT: DALBON, PASCAL
APPLICANT: LADAVIERE, LAURENT
APPLICANT: LACOUX, XAVIER
TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
TREATING AN HCV INFECTION
FILE REFERENCE: 103959
CURRENT APPLICATION NUMBER: US/10/367,677
PRIOR FILING DATE: 2003-02-19
PRIOR APPLICATION NUMBER: US/09/389,756
PRIOR FILING DATE: 1999-09-07
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 44
TYPE: PRT
ORGANISM: Hepatitis C virus
PUBLICATION INFORMATION:
AUTHORS: Ogata, N. et al.
TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
TITLE: Of Hepatitis Virus
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 3392-3396
DATE: 1991
RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match 84.1%; Score 127; DB 15; Length 44;
Best Local Similarity 92.3%; Pred. No. 2.2e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKNTIRRPQDVKPPGGVI 26
|||||
DB 4 PKPQRTKNTIRRPQDVKPPGGQI 29

RESULT 3

US-09-851-138-10
Sequence 10, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-851-138-10

Query Match 84.1%; Score 127; DB 10; Length 74;
Best Local Similarity 92.3%; Pred. No. 3.8e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKNTIRRPQDVKPPGGVI 26
|||||
DB 5 PKPQRTKNTIRRPQDVKPPGGQI 30

RESULT 4

US-09-758-308-1
Sequence 1, Application US/09758308
Patent No. US20020090607A1
GENERAL INFORMATION:
APPLICANT: HOWARD A. FIELDS AND YURY E. KHUYAKOV
TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C V
TITLE OF INVENTION: PROTEINS
FILE REFERENCE: 14114.0349U2
CURRENT APPLICATION NUMBER: US/09/758,308
CURRENT FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/092,339
PRIOR FILING DATE: 1999-07-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 91
TYPE: PRT
ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 84.1%; Score 127; DB 9; Length 91;
Best Local Similarity 92.3%; Pred. No. 4.8e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRTKNTIRRPQDVKPPGGVI 26
|||||
DB 5 PKPQRTKNTIRRPQDVKPPGGQI 30

;; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
;; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
;; TITLE OF INVENTION: applications thereof
;; FILE REFERENCE: B4809A - JAZ
;; CURRENT APPLICATION NUMBER: US/09/921.397
;; CURRENT FILING DATE: 2001-08-02
;; PRIOR APPLICATION NUMBER: EP 00402225.7
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 156
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 78
;; LENGTH: 113
;; TYPE: PRT
;; ORGANISM: Hepatitis C virus
US-09-921-397-78

Query Match 84.1%; Score 127; DB 10; Length 113;
Best Local Similarity 92.3%; Pred. No. 6.1e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Do 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
5 PKPQRTKTRNTIRRPQDVKPPGGGQI 30

RESULT 9

US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46

Query Match 84.1%; Score 127; DB 10; Length 137;
Best Local Similarity 92.3%; Pred. No. 7.5e-11;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
Db 5 PKPQRTKTRNTIRRPQDVKPPGGGQI 30

RESULT 10

US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO:
; AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-851-138-60

Query Match 84.1%; Score 127; DB 10; Length 138;
Best Local Similarity 92.3%; Pred. No. 7.6e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQRTKTRNTIRRPQDVKPPGGGVI 26
Db 5 PKPQRTKTRNTIRRPQDVKPPGGGQI 30

RESULT 11

US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-152

Query Match      84.1%; Score 127; DB 11; Length 166;
Best Local Similarity 92.3%; Pred. No. 9.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

      1 PKPQRTKRTNRRPDVKFPGGGVI 26
      ||||||| ||||||| |||||||
Db      5 PKPQRTKRTNRRPDVKFPGGGI 30

RESULT 12
US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-152

Query Match      84.1%; Score 127; DB 11; Length 166;
Best Local Similarity 92.3%; Pred. No. 9.3e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

      1 PKPQRTKRTNRRPDVKFPGGGVI 26
      ||||||| ||||||| |||||||
Db      5 PKPQRTKRTNRRPDVKFPGGGI 30

RESULT 13
US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
```

```
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-42

Query Match      84.1%; Score 127; DB 11; Length 169;
Best Local Similarity 92.3%; Pred. No. 9.5e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

      1 PKPQRTKRTNRRPDVKFPGGGVI 26
      ||||||| ||||||| |||||||
Db      5 PKPQRTKRTNRRPDVKFPGGGI 30

RESULT 14
US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-44

Query Match      84.1%; Score 127; DB 11; Length 169;
Best Local Similarity 92.3%; Pred. No. 9.5e-11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

      1 PKPQRTKRTNRRPDVKFPGGGVI 26
      ||||||| ||||||| |||||||
Db      5 PKPQRTKRTNRRPDVKFPGGGI 30

RESULT 15
US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-42

```

```

Query Match      84.1%; Score 127; DB 11; Length 169;
1st Local Similarity 92.3%; Pred. No. 9.5e-11;
atches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 PKPQRTKRNITRRPQDVKPPGGVI 26
Db      5 PKPQRTKRNITRRPQDVKPPGGVI 30

```

```

Search completed: August 7, 2003, 12:01:14
Job time : 15.3636 secs

```

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	147	95.5	3011	1	GNWVC3	genome polyprotein	
2	138	89.6	108	2	S41353	genome polyprotein	
3	138	88.6	108	2	S41355	genome polyprotein	
4	138	89.6	108	2	S41357	genome polyprotein	
5	138	89.6	108	2	S41348	genome polyprotein	
6	138	89.6	112	2	S41371	genome polyprotein	
7	138	89.6	112	2	S41341	genome polyprotein	
8	138	89.6	114	2	S41370	genome polyprotein	
9	138	89.6	114	2	S41369	genome polyprotein	
10	138	89.6	114	2	S41368	genome polyprotein	
11	138	89.6	115	2	S41342	genome polyprotein	
12	138	89.6	115	2	S41344	genome polyprotein	
13	138	89.6	115	2	S41350	genome polyprotein	
14	138	89.6	115	2	S41354	genome polyprotein	
15	138	89.6	115	2	S41345	genome polyprotein	
16	138	89.6	115	2	S41347	genome polyprotein	
17	138	89.6	115	2	S41343	genome polyprotein	
18	138	89.6	118	2	S41346	genome polyprotein	
19	138	89.6	369	2	S21471	genome polyprotein	
20	138	89.6	441	2	S12707	genome polyprotein	
21	138	89.6	513	2	PC1284	genome polyprotein	
22	138	89.6	520	2	JQ1925	genome polyprotein	
23	138	89.6	523	2	JQ1926	polyprotein - hepa	
24	138	89.6	550	2	JH0711	polyprotein - hepa	
25	138	89.6	782	2	S19876	genome polyprotein	
26	138	89.6	782	2	S18031	genome polyprotein	
27	138	89.6	782	2	S18032	genome polyprotein	
28	138	89.6	787	2	PK0677	hypothetical prote	
29	138	89.6	874	2	JQ0883	genome polyprotein	

```

Db      5 PKPQKNKRNTRRRPQDVKFPFGGGQIVG 32

RESULT 2
S41353
genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41353
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41353
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229456
A:Experimental source: genotype 2, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match      89.6%; Score 138; DB 2; Length 108;
Best Local Similarity 85.7%; Pred. No. 2.4e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQKNKRNTRRRPQDVKFPFGGGQIVG 28
      |||||: |||||: |||||: |||||: |||||:
Db      5 PKPQKTKRNTNRRPQDVKFPFGGGQIVG 32

RESULT 3
S41355
genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41355
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41355
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229458
A:Experimental source: genotype 2, N4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match      89.6%; Score 138; DB 2; Length 108;
Best Local Similarity 85.7%; Pred. No. 2.4e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQKNKRNTRRRPQDVKFPFGGGQIVG 28
      |||||: |||||: |||||: |||||: |||||:
Db      5 PKPQKTKRNTNRRPQDVKFPFGGGQIVG 32

RESULT 4
S41357
genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41357
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41357
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
A:Cross-references: EMBL:229474
A:Experimental source: genotype 5, N5
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>

Query Match      89.6%; Score 138; DB 2; Length 112;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQKNKRNTRRRPQDVKFPFGGGQIVG 28
      |||||: |||||: |||||: |||||: |||||:
Db      5 PKPQKTKRNTNRRPQDVKFPFGGGQIVG 32

RESULT 5
S41348
genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41348
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41348
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229451
A:Experimental source: genotype 1, N6
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match      89.6%; Score 138; DB 2; Length 108;
Best Local Similarity 85.7%; Pred. No. 2.4e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQKNKRNTRRRPQDVKFPFGGGQIVG 28
      |||||: |||||: |||||: |||||: |||||:
Db      5 PKPQKTKRNTNRRPQDVKFPFGGGQIVG 32

RESULT 6
S41371
genome polyprotein - hepatitis C virus (genotype 5, N5) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N5
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41371
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41371
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
A:Cross-references: EMBL:229474
A:Experimental source: genotype 5, N5
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>

Query Match      89.6%; Score 138; DB 2; Length 112;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQKNKRNTRRRPQDVKFPFGGGQIVG 28
      |||||: |||||: |||||: |||||: |||||:
Db      5 PKPQKTKRNTNRRPQDVKFPFGGGQIVG 32

```

Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVRFPGGGQIVG 28
||||: |||||||||
Db 5 PKPQKTKRNTNRRPDVRFPGGGQIVG 32

RESULT 7

S41341

genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 1, N1

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41341

R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

Accession: S41341

Molecule type: genomic RNA

A:Residues: 1-112 <VAN>

A:Cross-references: EMBL:Z29444; NID:g443850; PIDN:CAA82582.1; PID:g443851

A:Experimental source: genotype 1, N1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-112/Product: core protein #status predicted <NAT>

Query Match 89.6%; Score 138; DB 2; Length 112;

Best Local Similarity 85.7%; Pred. No. 2.5e-12;

Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVRFPGGGQIVG 28
||||: |||||||||
Db 5 PKPQKTKRNTNRRPDVRFPGGGQIVG 32

RESULT 8

S41370

genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 5, N4

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41370

R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

Accession: S41370

Molecule type: genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: EMBL:Z29473; NID:g443908; PIDN:CAA82611.1; PID:g443909

A:Experimental source: genotype 5, N4

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-114/Product: core protein #status predicted <NAT>

Query Match 89.6%; Score 138; DB 2; Length 114;

Best Local Similarity 85.7%; Pred. No. 2.5e-12;

Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVRFPGGGQIVG 28
||||: |||||||||
Db 5 PKPQKTKRNTNRRPDVRFPGGGQIVG 32

RESULT 9

S41369

genome polyprotein - hepatitis C virus (genotype 5, N3) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 5, N3

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41369

R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41369

A:Molecule type: genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: EMBL:Z29472; NID:g443906; PIDN:CAA82610.1; PID:g443907

A:Experimental source: genotype 5, N3

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-114/Product: core protein #status predicted <NAT>

Query Match 89.6%; Score 138; DB 2; Length 114;

Best Local Similarity 85.7%; Pred. No. 2.5e-12;

Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVRFPGGGQIVG 28
||||: |||||||||
Db 5 PKPQKTKRNTNRRPDVRFPGGGQIVG 32

RESULT 10

S41368

genome polyprotein - hepatitis C virus (genotype 5, N2) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 5, N2

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41368

R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41368

A:Molecule type: genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: EMBL:Z29471; NID:g443904; PIDN:CAA82609.1; PID:g443905

A:Experimental source: genotype 5, N2

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-114/Product: core protein #status predicted <NAT>

Query Match 89.6%; Score 138; DB 2; Length 114;

Best Local Similarity 85.7%; Pred. No. 2.5e-12;

Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVRFPGGGQIVG 28
||||: |||||||||
Db 5 PKPQKTKRNTNRRPDVRFPGGGQIVG 32

RESULT 11

S41342

genome polyprotein - hepatitis C virus (genotype 1, N10) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 1, N10

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41342

R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41342

A:Molecule type: genomic RNA

A:Residues: 1-115 <VAN>

A:Cross-references: EMBL:Z29445; NID:g443852; PIDN:CAA82583.1; PID:g443853

A:Experimental source: genotype 1, N10

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.6%; Score 138; DB 2; Length 115;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPQDVKPPGGQIVG 28
||||: |||||||||
DB 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32

RESULT 12

S41344

genome polyprotein - hepatitis C virus (genotype 1, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41344
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

A:Accession: S41344
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229447; NID:g443856; PIDN:CAA82585.1; PID:g443857
A:Experimental source: genotype 1, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.6%; Score 138; DB 2; Length 115;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPQDVKPPGGQIVG 28
||||: |||||||||
DB 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32

RESULT 13

S41350

genome polyprotein - hepatitis C virus (genotype 1, N8) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N8
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41350

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

A:Accession: S41350
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229453; NID:g443868; PIDN:CAA82591.1; PID:g443869
A:Experimental source: genotype 1, N8
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.6%; Score 138; DB 2; Length 115;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPQDVKPPGGQIVG 28
||||: |||||||||
DB 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32

RESULT 14

S41354

genome polyprotein - hepatitis C virus (genotype 2, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41354
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

A:Accession: S41354
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229457
A:Experimental source: genotype 2, N3

C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.6%; Score 138; DB 2; Length 115;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPQDVKPPGGQIVG 28
||||: |||||||||
DB 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32

RESULT 15

S41345

genome polyprotein - hepatitis C virus (genotype 1, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41345
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341

A:Accession: S41345
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229448; NID:g443858; PIDN:CAA82586.1; PID:g443859

A:Experimental source: genotype 1, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.6%; Score 138; DB 2; Length 115;
Best Local Similarity 85.7%; Pred. No. 2.5e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPQDVKPPGGQIVG 28
||||: |||||||||
DB 5 PKPQKTKRNTNRRPQDVKPPGGQIVG 32

Search completed: August 7, 2003, 11:21:48
Job time: 9.54545 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146A-29

Perfect score: 154

Sequence: 1 PKPQRNQRNRRPDQVKGPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	95.5	3011	1 POLG_HCV1	P26664 h genome po
2	138	89.6	513	1 POLG_HCVJ2	P27959 hepatitis c
3	138	89.6	520	1 POLG_HCVH4	Q01404 hepatitis c
4	138	89.6	520	1 POLG_HCVHK	Q01403 hepatitis c
5	138	89.6	737	1 POLG_HCVJ5	P27960 hepatitis c
6	138	89.6	737	1 POLG_HCVJ7	P27961 hepatitis c
7	138	89.6	3010	1 POLG_HCVBK	P26663 h genome po
8	138	89.6	3011	1 POLG_HCVJA	P26662 h genome po
9	138	89.6	3011	1 POLG_HCVH	P27958 h genome po
10	138	89.6	3033	1 POLG_HCVJ6	P26660 h genome po
11	138	89.6	3033	1 POLG_HCVJ8	P26661 h genome po
12	131	85.1	3010	1 POLG_HCVTW	P29846 h genome po
13	130	84.4	3010	1 POLG_HCVJT	Q00269 h genome po
14	54.5	35.4	586	1 2A5D_RABIT	Q28653 o serine/th
15	54.5	35.4	602	1 2A5D_HUMAN	Q14738 h serine/th
16	53	34.4	282	1 RK4_TOBAC	O80361 nicotiana t
17	52.5	34.1	303	1 YB1_XENLA	P21573 xenopus lae
18	50	32.5	440	1 YD49_YEAST	Q02354 saccharomyc
19	50	32.5	979	1 PUT3_YEAST	P25502 saccharomyc
20	49	31.8	950	1 IF2_LALIC	Q9X764 lactococcus
21	49	31.8	1960	1 TF20_HUMAN	Q9UGU0 homo sapien
22	49	31.8	1983	1 TF20_MOUSE	Q9EPQ8 mus musculu
23	48	31.2	240	1 DAPB_STAAM	O99U88 staphylococ
24	48	31.2	240	1 DAPB_STAAM	O9EX11 staphylococ
25	48	31.2	240	1 DAPB_STAAM	O8NW54 staphylococ
26	48	31.2	461	1 SYN_FUSNN	Q8RH70 fusbacteri
27	48	31.2	556	1 PDPK_HUMAN	O15530 homo sapien
28	48	31.2	825	1 ABF1_STRCX	P82593 schistosomyc
29	47.5	30.8	407	1 PUSA_SCHPO	O59721 schizosacch
30	47	30.5	339	1 CSP_PLABE	P06915 plasmodium
31	47	30.5	347	1 CSP_PLABA	P23093 plasmodium
32	47	30.5	436	1 GDF6_BOVIN	P55106 bos taurus
33	47	30.5	485	1 GLC2_SOYBN	P04405 glycine max

RESULT 1

POLG_HCV1

ID POLG_HCV1 STANDARD; PRT; 3011 AA.

AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome Polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate 1) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;
 RA "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; M62321; AAA45676.1;
 DR PIR; A39166; GNMVC3.
 DR PDB; 1A1V; 16-FEB-99.
 DR PDB; 1HEI; 25-NOV-98.
 DR MEROPS; S29.001;
 DR MEROPS; U39.001;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.

ALIGNMENTS

P10723 brugia mala
 P51816 homo sapien
 P21711 xenopus lae
 P08855 oryctolagus
 P49428 pichia past
 O9C2X5 mus musculu
 Q922A0 mus musculu
 O55173 rattus norv
 Q13625 homo sapien
 P18123 zea mays (m
 Q04089 saccharomyc
 P05674 venezuelan

1 SYN_BRUMA
 1 FMR2_HUMAN
 1 MIX1_XENLA
 1 ICAL_RABIT
 1 UBCX_PICPA
 1 P111_MOUSE
 1 PDPK_MOUSE
 1 PDPK_RAT
 1 P532_HUMAN
 1 CAT3_MAIZE
 1 DOT1_YEAST
 1 POLS_EEUV8

34 47 30.5 548
 35 47 30.5 1311
 36 46.5 30.2 377
 37 46.5 30.2 718
 38 46 29.9 204
 39 46 29.9 332
 40 46 29.9 559
 41 46 29.9 559
 42 46 29.9 1005
 43 45.5 29.5 496
 44 45.5 29.5 582
 45 45.5 29.5 1254

DR	InterPro; IPR002521; HCV_core.	
DR	InterPro; IPR002519; HCV_env.	
DR	InterPro; IPR002531; HCV_NS1.	
DR	InterPro; IPR002518; HCV_NS2.	
DR	InterPro; IPR004109; HCV_NS3.	
DR	InterPro; IPR000745; HCV_NS4a.	
DR	InterPro; IPR0011490; HCV_NS4b.	
DR	InterPro; IPR002868; HCV_NS5a.	
DR	InterPro; IPR002166; HCV_RdRp.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_Psvir.	
DR	Pfam; PF01543; HCV_capsid; 1.	
DR	Pfam; PF01542; HCV_core; 1.	
DR	Pfam; PF01539; HCV_env; 1.	
DR	Pfam; PF01560; HCV_NS1; 1.	
DR	Pfam; PF01538; HCV_NS2; 1.	
DR	Pfam; PF02907; HCV_NS3; 1.	
DR	Pfam; PF01006; HCV_NS4a; 1.	
DR	Pfam; PF01001; HCV_NS4b; 1.	
DR	Pfam; PF01506; HCV_NS5a; 1.	
DR	Pfam; PF00271; helicase_C; 1.	
DR	Pfam; PF00998; Viral_RdRp; 1.	
DR	ProDom; PD186062; HCV_NS1; 1.	
DR	SMART; SM00487; DEXdc; 1.	
DR	Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolyase; Serine protease; 3D-structure.	
FT	INIT_MET	1
FT	CHAIN	1
FT	CHAIN	115
FT	CHAIN	191
FT	CHAIN	383
FT	CHAIN	729
FT	CHAIN	1006
FT	CHAIN	1615
FT	CHAIN	1862
FT	CHAIN	2013
FT	CHAIN	3011
FT	CHAIN	347
FT	CHAIN	369
FT	TRANSMEM	1083
FT	ACT_SITE	1107
FT	ACT_SITE	1107
FT	ACT_SITE	1165
FT	ACT_SITE	1230
FT	NP_BIND	1316
FT	SITE	1319
FT	CARBOHYD	196
FT	CARBOHYD	209
FT	CARBOHYD	234
FT	CARBOHYD	305
FT	CARBOHYD	417
FT	CARBOHYD	423
FT	CARBOHYD	430
FT	CARBOHYD	448
FT	CARBOHYD	476
FT	CARBOHYD	532
FT	CARBOHYD	540
FT	CARBOHYD	556
FT	CARBOHYD	576
FT	CARBOHYD	623
FT	CARBOHYD	645
FT	CARBOHYD	2041
FT	CARBOHYD	2077
FT	CARBOHYD	2240
FT	CARBOHYD	2364
FT	CARBOHYD	2789
FT	CARBOHYD	3011
FT	SEQUENCE	327197
FT	AA;	65F8C947FCE5AF9 CRC64;
FT	Query Match	95.5%;
FT	Best Local Similarity	92.9%;
FT	Matches	26; Conservative
FT	Mismatches	2;
FT	Indels	0;
FT	Gaps	0;
FT	1 PKPKQRNQRNTNRPDVVFPGGGQIVG	28

```

FT NON_TER      520
SQ SEQUENCE     520 AA; 56499 MW; AAL35246CF20D525 CRC64;

Query Match          89.68; Score 138; DB 1; Length 520;
Best Local Similarity 85.7%; Pred. No. 1.6e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRKNQRNRRPQDVKEPPGGQIVG 28
    ||||| :|||||
Db 5 PKPORKTRNTNRNPQDVKEPPGGQIVG 32

RESULT 4
POLG_HCVHKH
ID ID POLG_HCVHKH STANDARD; PRT; 520 AA.
AC O01403;
DT DT 01-JUL-1993 (Rel. 26, Created)
DE DE 01-JUL-1993 (Rel. 26, Last sequence update)
DI DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE DE (GP68) (GP70) (NS1)] (Fragment).
OS OS Hepatitis C virus (isolate HCV-KF) (HCV).
OC OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
OC Hepacivirus.
NC NC NCBI_TaxID=31644;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RA RA MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL RL J. Gen. Virol. 73:2725-2729(1992).
CC CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC EMBL; D10687; BAA01529.1; -.
DR DR PIR; JQ1925; JQ1925.
DR DR InterPro; IPR002522; HCV_capsid.
DR DR InterPro; IPR002521; HCV_core.
DR DR InterPro; IPR002519; HCV_env.
DR DR InterPro; IPR002531; HCV_NS1.
DR DR Pfam; PF01543; HCV_capsid; 1.
DR DR Pfam; PF01542; HCV_core; 1.
DR DR Pfam; PF01539; HCV_env; 1.
DR DR Pfam; PF01560; HCV_NS1; 1.
DR DR ProDom; PD186062; HCV_NS1; 1.
DR DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
DR Transmembrane; Nonstructural protein.
KW KW INIT_MET 1 1
FT FT REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115
FT CHAIN 116 191
FT FT MATRIX PROTEIN C (POTENTIAL).
FT CHAIN 192 383
FT FT MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 >520
FT FT NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT FT BY SIMILARITY.
FT FT TRANSMEM 347 369
FT FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT FT CARBOHYD 196 196
FT FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT FT CARBOHYD 209 209
FT FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT FT CARBOHYD 233 233
FT FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT FT CARBOHYD 234 234
FT FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT FT CARBOHYD 305 305
FT FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT FT CARBOHYD 418 418

```

FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match
 Best Local Similarity 89.6%; Score 138; DB 1; Length 520;
 Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPKRNORNTNRPPQDVKPPGGQIVG 28
 Db 5 PKPQRTKRNTRPPQDVKPPGGQIVG 32

RESULT 5
 POLG_HCVJ5
 ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
 AC P27960;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix proteins NS1 and NS2] (fragment).
 CC Hepatitis C virus (isolate HC-J5) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT Full-length sequence of a hepatitis C virus genome having poor
 RT genotypes*.
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D10075; BAA00969.1;
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NSI.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_core; 1.
 CC Pfam: PF01538; HCV_NSI; 1.
 CC Pfam: PF01560; HCV_NSI; 1.
 CC ProDom: PD186062; HCV_NSI; 1.
 CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 CC Transmembrane; Nonstructural protein.
 CC INIT_MET 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 >737
 CC NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 CC NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT TRANSMEM 347 369 POTENTIAL.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 737 AA; 81207 MW; 3AF69D82AD501B1 CRC64;

Query Match
 Best Local Similarity 89.6%; Score 138; DB 1; Length 737;
 Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPKRNORNTNRPPQDVKPPGGQIVG 28
 Db 5 PKPQRTKRNTRPPQDVKPPGGQIVG 32

RESULT 6
 POLG_HCVJ7
 ID POLG_HCVJ7 STANDARD; PRT; 737 AA.
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix proteins NS1 and NS2] (fragment).
 CC Hepatitis C virus (isolate HC-J7) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT Full-length sequence of a hepatitis C virus genome having poor
 RT genotypes*.
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D10077; BAA00971.1;
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_core; 1.
 CC Pfam: PF01538; HCV_NSI; 1.
 CC ProDom: PD186062; HCV_NSI; 1.
 CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 CC Transmembrane; Nonstructural protein.
 CC INIT_MET 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 733
 CC CHAIN 734 >737
 CC NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 CC NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 ProDom: PD186062; HCV_NS1; 1.
 Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 Transmembrane; Nonstructural protein.
 INIT_MET 1 1
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 733
 CHAIN 734 >737
 TRANSEM 347 369
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 233 233
 CARBOHYD 299 299
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 477 477
 CARBOHYD 534 534
 CARBOHYD 542 542
 CARBOHYD 558 558
 CARBOHYD 578 578
 CARBOHYD 627 627
 CARBOHYD 649 649
 NON_TER 737 737
 SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 89.6%; Score 138; DB 1; Length 737;
 Best Local Similarity 85.7%; Pred. No. 2.3e-12;
 Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPKRNRQNTNRPPQVDPKPGGQIVG 28
 DB 5 PKPQKTKRNTNRPPQVDPKPGGQIVG 32

RESULT 7
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 D 01-AUG-1992 (Rel. 23, Last sequence update)
 D 15-SEP-2003 (Rel. 42, Last annotation update)
 L Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
 CC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RN MEDLINE-91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 from human carriers."
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RN SEQUENCE OF 1487-1500.
 RP MEDLINE-96235224; PubMed=8647104;
 RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetseky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits

phosphorylation mediated by cAMP-dependent protein kinase.";
 Eur. J. Biochem. 237:611-618(1996).
 [3]
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE-97015088; PubMed=8861916;
 RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 Moomaw E.W., Adachi T., Hostomska Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 trypsin-like fold and a structural zinc binding site.";
 Cell 87:331-342(1996).
 [4]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE-98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL; M58335; AAA72945.1;
 DR PIR; A38465; GNVVTC.
 DR PDB; 1A10; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QUV; 26-JUN-00.
 DR PDB; 8OHM; 20-APR-99.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_psvir.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 Pfam; PF01538; HCV_NS2; 1.
 Pfam; PF02907; HCV_NS3; 1.
 Pfam; PF01006; HCV_NS4a; 1.

DR HSP: P26663; LJXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR004109; HCV NS3.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV NS5b.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_Ps.
 DR InterPro: IPR007094; RNA_pol_Psvir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT VP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794f46DB185 CRC64;
 Query Match 89.6%; Score 138; DB 1; Length 3010;
 Best Local Similarity 85.7%; Pred. No. 1,1e-11;
 Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVYKFGGGQIVG 28
 DB 5 PKPQKTRKNTNRRPDVYKFGGGQIVG 32
 RESULT 9
 POLG_HCVH STANDARD; PRT: 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P44); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=111108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-92052256; PubMed-1658800;
 RX Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE-97331322; PubMed-9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE-98154321; PubMed-9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -|- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -|- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -|- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -|- FUNCTION: NS5B SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -|- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -|- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -|- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -|- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

..... entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CCC

	EURL; M67463; AAA45534.1;	-	
DR	PIR; A36814; GNMVCH.		
DR	PDB; IHEI; 25-NOV-98.		
DR	PDB; IAIY; 16-FEB-99.		
DR	PDB; IAIR; 17-JUN-98.		
DR	MEROPS; S29_001; -		
DR	MEROPS; U39_001; -		
DR	TRANSFAC; T04155; -		
DR	InterPro; IPRO01410; DEAD.		
DR	InterPro; IPRO02522; HCV_capsid.		
DR	InterPro; IPRO02521; HCV_core.		
DR	InterPro; IPRO02519; HCV_env.		
DR	InterPro; IPRO02531; HCV_NS1.		
DR	InterPro; IPRO02518; HCV_NS2.		
DR	InterPro; IPRO04109; HCV_NS3.		
DR	InterPro; IPRO00745; HCV_NS4a.		
DR	InterPro; IPRO01490; HCV_NS4b.		
DR	InterPro; IPRO02868; HCV_NS5a.		
DR	InterPro; IPRO02166; HCV_RdRp.		
DR	InterPro; IPRO01650; Helicase_C.		
DR	InterPro; IPRO07095; RNA_pol_DS_Psr.		
DR	InterPro; IPRO07094; RNA_pol_Psvir.		
DR	Pfam; PF01543; HCV_capsid; 1.		
DR	Pfam; PF01542; HCV_core; 1.		
DR	Pfam; PF01539; HCV Env; 1.		
DR	Pfam; PF01560; HCV_NS1; 1.		
DR	Pfam; PF01538; HCV_NS2; 1.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF01006; HCV_NS4a; 1.		
DR	Pfam; PF01001; HCV_NS4b; 1.		
DR	Pfam; PF01506; HCV_NS5a; 1.		
DR	Pfam; PF00271; helicase_C; 1.		
DR	Pfam; PF00998; Viral_RdRP; 1.		
DR	PRODom; PD186062; HCV_NS1; 1.		
DR	SMART; SM00487; DEXDC; 1.		
KW	Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;		
KW	Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;		
KW	Transmembrane; Nonstructural protein; Hydrolase; Serine protease;		
F	3D-structure.		
FT	INIT_MET 1 1		
FT	CHAIN 1 191		REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.
FT	CHAIN 192 383		CAPSID PROTEIN C.
FT	CHAIN 384 746		ENVELOPE GLYCOPROTEIN E1.
FT	CHAIN 747 809		ENVELOPE GLYCOPROTEIN E2.
FT	CHAIN 810 1026		PROTEIN P7.
FT	CHAIN 1027 1657		NONSTRUCTURAL PROTEIN NS2.
FT	CHAIN 1658 1711		PROTEASE/HELICASE NS3.
FT	CHAIN 1712 1972		NONSTRUCTURAL PROTEIN NS4A.
F	CHAIN 1973 2420		NONSTRUCTURAL PROTEIN NS4B.
F	CHAIN 2421 3011		NONSTRUCTURAL PROTEIN NS5A.
FT	CHAIN TRANSMEM 347 369		NONSTRUCTURAL PROTEIN NS5B.
FT	ACT_SITE 1083 1083		POTENTIAL.
FT	ACT_SITE 1107 1107		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 1165 1165		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	NP_BIND 1230 1237		ATP (POTENTIAL).
FT	SITE 1316 1319		DECK BOX.
FT	CARBHYD 196 196		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 209 209		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 234 234		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 305 305		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 417 417		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 423 423		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 430 430		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 448 448		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 476 476		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 532 532		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 540 540		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 556 556		N-LINKED (GLCNAC..) (POTENTIAL).
FT	CARBHYD 576 576		N-LINKED (GLCNAC..) (POTENTIAL).

	623	623	N-LINKED (GLCNAC. .)	(POTENTIAL) . .
FT CARBOHYD	623			
FT CARBOHYD	645		N-LINKED (GLCNAC. .)	(POTENTIAL) . .
FT STRAND	1224	1226		

SQ	SEQUENCE	3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Query Match	89.6%	Score 138;	DB 1;	Length 3011;
Best Local Similarity	85.7%	Prod N		

Matches	Conservative	Mismatches	Indels	Gaps
24	3	1	0	0

QY 1 PKPQKRNQRTNRRPQDVKFPGGQIVG 28
 ||||: :|||||:|||||
 Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 10
POLG_HCVJ6

ID	POLG_HCVJ6	STANDARD;	PRT; 3033 AA.
----	------------	-----------	---------------

	Pfam:	PF01506;	HCV_NS5a; 1.
DR	Pfam:	PF00271;	Helicase_C; 1.
DR	Pfam:	PF00998;	Viral_RdRP; 1.
DR	PfDom;	PD186062;	HCV_NS1; 1.
DR	SMART;	SM00487;	DExDc; 1.
KW	Polyprotein:	Glycoprotein;	Transferase; RNA-directed RNA polymerase;
KV	Core protein:	Coat protein;	Envelope protein; Helicase; ATP-binding;
KW	Transmembrane;	Nonstructural	protein; Hydrolase; Serine protease.
FT	INIT_MET	1	REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.
FT	CHAIN	1	CAPSID PROTEIN C (POTENTIAL).
FT	CHAIN	116	MATRIX PROTEIN (POTENTIAL).
FT	CHAIN	191	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	CHAIN	192	NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT	CHAIN	384	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT	CHAIN	734	PROTEASE/Helicase NS3 (POTENTIAL).
FT	CHAIN	1010	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT	CHAIN	1619	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT	CHAIN	1620	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT	CHAIN	1866	POTENTIAL.
FT	CHAIN	1867	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CHAIN	2017	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CHAIN	2018	ATP (POTENTIAL).
FT	TRANSMEM	347	DECH BOX.
FT	ACT_SITE	1087	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	ACT_SITE	1087	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	ACT_SITE	1111	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	ACT_SITE	1111	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	ACT_SITE	1169	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	NP_BIND	1234	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	NP_BIND	1234	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SITE	1320	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SITE	1320	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	196	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	209	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	234	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	305	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	417	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	423	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	430	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	448	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	477	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	477	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	534	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	542	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	558	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	578	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	627	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	649	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1091	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2038	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2811	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2811	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SEQUENCE	3033 AA;	F957F5C1A273BE9E CRC64;
SQL			
<hr/>			
	Query Match	89.6%;	Score 138; DB 1; Length 3033;
	Best Local Similarity	85.7%;	Pred. No. 1.le-11;
	Matches	24; Conservative	3; Mismatches 1; Indels 0; Gaps
Oy	1	PKPQRNQNTNRRPODVKFGGQIVG 28	
		:	
Db	5	PKPQRKTENTNRRPODVKFGGQIVG 32	
<hr/>			
	RESULT 11		
	POLG_HCVJ8		
ID	POLG-HCVJ8	STANDARD;	PRT; 3033 AA..
AD	P26661;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]. Hepatitis C virus (isolate HC-J8) (HCV). OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; OC Hepacivirus. NCBI_TaxID=11115; XN [1]		

SEQUENCE FROM N.A.
MEDLINE=92230232; PubMed=1314459;
Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
Fukuda S., Tsuda F., Mishi S.
"Full-length sequence of a hepatitis C virus genome having poor
homology to reported isolates: comparative study of four distinct
genotypes.";
Virology 188:331-341(1992).
-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
-!- NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the
precursor polypeptide, commonly with Asp or Glu in the p6
position, Cys or Thr in p1 and Ser or Ala in p1'.
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate +
[RNA](N).
-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
send an email to license@isb-sib.ch).

DR EMBL; D10988; BAA01761.1; -
DR PIR: A40250; C10988

DR HSSP; P27958; THEI.
DR MEROPS. 200 241

MEROPS; 529.001; -
MEROPS; 539.001; -

InterPro: IPR001410; DEAD, uc001luc.1
InterPro: IPR002522; uc001luc.1

```

InterPro: IPR002521; HCV_core.
InterPro: IPR002522; hcv_capsid.
InterPro: IPR002523; HCV_N.

```

InterPro; IPR002531; HCV_NSI.
InterPro; IPR002531; HCV_NSI.

```

InterPro; IPR002518; HCV_NS2.
InterPro; IPR004109; HCV_NS3

```

InterPro; IPR000745; HCV_NS4a.
InterPro; IPR001490; HCV_NS4a.

InterPro: IPR002868; HCV_NS5a, InterPro: IPR000150; HCV_NS4b, InterPro: IPR000150

InterPro: IPR007095; RNA pol DS ps

interfiro; IPR007094; RNA_pol_PSvir.
Pfam; PF01543; HCV capsid.1

Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_core; 1.

Pfam: PF01560; HCV_NSI; 1.
Pfam: PF01538; HCV_NSI; 1.

fam; PF02907; HCV_NS3; 1;
 fam; PF02907; HCV_NS3; 1;

.uni; PF01006; HCV_NS4a; 1.
 .fam; PF01001; HCV_NS4b; 1.
 .ns

Pfam; PF01506; HCV_NS5a; 1.
Pfam; PF00998; Viral protease; 1.

ProdDom; PD186062; HCV_NSI; 1.
SMART; SM00497; 1.

Polyprotein; Glycoprotein; Transferrin. See 1.

Core protein; Coat protein; Envelope protein; Helicase; ATP-binding site; RNA polymerase; RNA-directed RNA polymerase; Transmembrane; Nonstructural protein

INIT_MET
1
protein;Hydrolase; Serine protease.
REMOVED FROM CAPSID PROTEIN C

CHAIN	1	115
CHAIN	116	101

CHAIN	RESIDUE NO.	FUNCTIONAL GROUP
CHAIN 1	191	MATRIX PROTEIN (POTENTIAL).
CHAIN 2	192	MATRIX PROTEIN (POTENTIAL).
CHAIN 3	383	MAJOR ENVELOPE PROTEIN (POTENTIAL).
CHAIN 4	384	MAJOR ENVELOPE PROTEIN (POTENTIAL).

CHAIN	384	733	734	1010
MEMBRANE ENVELOPE PROTEIN E (POTENTIAL).				
NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).				
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).				

CHAIN	1011	1619	1620	1866
STRUCTURAL PROTEIN NS2 (POTENTIAL)				
PROTEASE/HELICASE NS3 (POTENTIAL)				

	CHAIN	
- - -	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL)	1867
- - -	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL)	1900
- - -	NONSTRUCTURAL PROTEIN NS4C (POTENTIAL)	1930
- - -	NONSTRUCTURAL PROTEIN NS4D (POTENTIAL)	1967
- - -	NONSTRUCTURAL PROTEIN NS4E (POTENTIAL)	1999
- - -	NONSTRUCTURAL PROTEIN NS4F (POTENTIAL)	2017

PROTEIN NS4B (POTENTIAL).

FT	CHAIN	2018	3033	RNA-DIRECTED RNA POLYMERASE (POTENTIAL)
FT	TRANSMEM	347	369	POTENTIAL.
FT	ACT_SITE	1087	1087	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1111	1111	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1169	1169	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	NP_BIND	1234	1241	ATP (POTENTIAL).
FT	SITE	1320	1323	DECH BOX.
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	233	233	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	299	299	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	477	477	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	534	534	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	542	542	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	558	558	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	578	578	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	627	627	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	649	649	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1091	1091	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2038	2038	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2359	2359	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2811	2811	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3033	3033	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	3033	3033	AA; 330177 MW; IAL173E7E3381FDIA CRC64;
Query Match				89.6%; Score 138; DB 1; Length 3033;
Best Local Similarity				85.7%; Pred. No. 1.1e-11;
Matches				24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY	1	PKPQRNORNTNRRPQDKVFGGGQIVG	28	
Db	5	PKPQRKTRNTNRRPQDKVFGGGQIVG	32	
RESULT 12				
POLG_HCVTW				
ID	POLG_HCVTW	STANDARD;	PRT;	3010 AA.
AC	P29846;			
DT	01-APR-1993	(Rel. 25, Created)		
DT	01-APR-1993	(Rel. 25, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Envelope glycoprotein [Contains: Capsid protein C (Core protein) (P22);			
DE	(GP68), (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21);			
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)			
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein			
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein			
DE	NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
OS	Hepatitis C virus (Isolate Taiwan) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OX	Hepacivirus.			
NCBI_TaxID	31645;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-92230206; PubMed-1314449;			
RA	Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;			
RT	"The Taiwanese Hepatitis C virus genome: sequence determination and			
RL	mapping the 5' termini of viral genomic and antigenomic RNA.";			
CC	Virology 188:102-113(1992).			
CC	-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE			
CC	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral			
CC	precursor polyprotein, commonly with Asp or Glu in the P6			
CC	position, Cys or Thr in P1 and Ser or Ala in P1.			
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +			
CC	[RNA](N).			
CC	-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A			
CC	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:			

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CCR PIR: M84754; -; NOT_ANNOTATED_CDS.
 CCR PIR: A40244; GNVVTV.
 CCR PDB: 1N64; 25-FEB-03.
 CCR PDB: 1NS3; 08-APR-98.
 CCR MEROPS: S29.001; -.
 CCR InterPro: IPR001410; DEAD.
 CCR InterPro: IPR002522; HCV_capsid.
 CCR InterPro: IPR002521; HCV_core.
 CCR InterPro: IPR002519; HCV_env.
 CCR InterPro: IPR002531; HCV_NS1.
 CCR InterPro: IPR002518; HCV_NS2.
 CCR InterPro: IPR004109; HCV_NS3.
 CCR InterPro: IPR000745; HCV_NS4a.
 CCR InterPro: IPR001490; HCV_NS4b.
 CCR InterPro: IPR002868; HCV_NS5a.
 CCR InterPro: IPR002166; HCV_RdRp.
 CCR InterPro: IPR007095; RNA_pol_DS_PS.
 CCR InterPro: IPR007094; RNA_pol_PSVir.
 CCR Pfam: PF01543; HCV_capsid; 1.
 CCR Pfam: PF01542; HCV_core; 1.
 CCR Pfam: PF01539; HCV_NS1; 1.
 CCR Pfam: PF01560; HCV_NS2; 1.
 CCR Pfam: PF01538; HCV_NS3; 1.
 CCR Pfam: PF02907; HCV_NS4a; 1.
 CCR Pfam: PF01006; HCV_NS4b; 1.
 CCR Pfam: PF01001; HCV_NS5a; 1.
 CCR Pfam: PF00271; helicase_C; 1.
 CCR Pfam: PF00998; Viral_RdRp; 1.
 CCR ProDom: PD186062; HCV_NS1; 1.
 CCR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 F INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 F CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423

FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDE215 CRC64;
 Query Match 85.1%; Score 131; DB 1; Length 3010;
 Best Local Similarity 85.2%; Pred. No. 1.2e-10;
 Matches 23; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 KPQKRNQNTNRPPQDVKPPGGQIVG 28
 DB 6 KPQKTKRNTNRPPQDVKPPGGQIVG 32
 RESULT 13
 POLG_HCVJT
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.;
 RL Virus Res. 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC {RNA}(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; D11168; BAA01943.1; .
 DR PIR; A45573; A45573.
 DR PDB; 1A10; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR MEROPS; S29.001; .
 DR MEROPS; U39.001; .
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002866; HCV_NS5b.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_Psvir.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RGRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.
 DR INIT_MET 1 1
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN TRANSMEM 347 369
 ACT_SITE 1083 1083
 ACT_SITE 1107 1107
 ACT_SITE 1165 1165
 NP_BIND 1230 1237
 SITE 1316 1319
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 250 250
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 532 532
 CARBOHYD 540 540
 CARBOHYD 556 556
 CARBOHYD 576 576
 CARBOHYD 623 623
 CARBOHYD 645 645
 CARBOHYD 2041 2041
 CARBOHYD 2077 2077
 CARBOHYD 2240 2240
 CARBOHYD 2529 2529
 CARBOHYD 2788 2788

SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 84.4%; Score 130; DB 1; Length 3010;
 Best Local Similarity 82.1%; Pred. NO. 1.6e-10;
 Matches 23; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 PPKQKRNQNTNRPPQDVKFPGGQIVG 28
 DB 5 PKPQKRNQNTNRPPQDVKFPGGQIVG 32

RESULT 14

2A5D.RABIT STANDARD; PRT; 586 AA.
 AC Q28653; Q28655;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine protein phosphatase 2A, 56 kDa regulatory subunit,
 DE delta isoform (PP2A, B subunit, B' delta isoform) (PP2A, B subunit,
 DE B56 delta isoform) (PP2A, B subunit, B' delta isoform) (PP2A, B
 DE subunit, R5 delta isoform) (PP2A, B subunit, B' gamma).
 OS PP2R5D.
 GN Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID-9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-New Zealand; TISSUE-Brain, and Skeletal muscle;
 RX MEDLINE-96161994; PubMed-8576224;
 RA Gortos C.; Zolnierowicz S.; Bako E.; Durbin S.D.; Depaoli-Roach A.A.;
 RT "High complexity in the expression of the B' subunit of protein
 RT phosphatase 2A0. Evidence for the existence of at least seven novel
 RT isoforms.";
 RL J. Biol. Chem. 271:2578-2588(1996).
 CC -!- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
 CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
 CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
 CC COMPARTMENT.
 CC -!- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
 CC COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
 CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A). THAT ASSOCIATES
 CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
 CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
 CC B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
 CC FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
 CC AND CELL SIGNALING MOLECULES.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- TISSUE SPECIFICITY: Highly expressed in brain.
 CC -!- SIMILARITY: Belongs to the phosphatase 2A regulatory subunit B
 CC family.
 CC -!- CAUTION: NOMENCLATURE USED IN REF.1 REFERS TO PP2A B SUBUNIT
 CC B' GAMMA ISOFORM, WHICH IS CITED AS PP2A B SUBUNIT DELTA-PR61
 CC ISOFORM IN LATER PUBLICATIONS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U38193; AAC48532.1;
 CC EMBL; U38195; AAC48534.1;
 CC InterPro; IPR002554; B56.
 CC Pfam; PF01603; B56; 1.
 CC Multigene family; Nuclear:protein; Repeat.
 CC DOMAIN 21 36 8 X 2 AA APPROXIMATE TANDEM REPEATS OF Q-
 CC P.
 CC SH3 BINDING, CLASS I (POTENTIAL).
 CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC -----
 CC DOMAIN 507 514
 CC DOMAIN 532 549

Matches 13; Conservative 9; Mismatches 5; Indels 11; Gaps 2;
Qy 1 PKPKRNO-RNTNRPOD-----VKFPGGGQIV 27
Db 44 POPAQSOPPSSNRPSNSTPPTQLSKIRYGGPQIV 81

Search completed: August 7, 2003, 11:20:02
Job time : 5.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-29
Perfect score: 154
Sequence: 1 PKQKRNQRTNRPQDVKPPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues 830525
1 number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	415	12 P87751	P87751 hepatitis c
2	154	100.0	3013	12 O92530	O92530 hepatitis c
3	146	94.8	191	12 O8V7Y2	O8V7Y2 hepatitis c
4	142	92.2	191	12 O68137	O68137 hepatitis c
5	141	91.6	103	12 O9E983	O9E983 hepatitis c
6	141	91.6	3010	12 O9J3H8	O9J3H8 hepatitis c
7	141	91.6	3010	12 O9QIY9	O9QIY9 hepatitis c
8	140	90.9	191	12 O8V7V3	O8V7V3 hepatitis c
9	140	90.9	191	12 O8V7W0	O8V7W0 hepatitis c
10	140	90.9	191	12 O8V7W8	O8V7W8 hepatitis c
11	140	90.9	191	12 O8V7W5	O8V7W5 hepatitis c
12	139	90.3	191	12 O8JWL7	O8JWL7 hepatitis c
13	138	89.6	45	12 O68311	O68311 hepatitis c
14	138	89.6	45	12 O68306	O68306 hepatitis c
15	138	89.6	46	12 O68309	O68309 hepatitis c
16	138	89.6	60	12 O8JYR9	O8JYR9 hepatitis c

17	138	89.6	61	12 O8JYR8	O8JYR8 hepatitis c
18	138	89.6	61	12 O8JYR5	O8JYR5 hepatitis c
19	138	89.6	61	12 O8JYR7	O8JYR7 hepatitis c
20	138	89.6	62	12 O8JYR6	O8JYR6 hepatitis c
21	138	89.6	74	12 O68712	O68712 hepatitis c
22	138	89.6	74	12 O68682	O68682 hepatitis c
23	138	89.6	74	12 O68708	O68708 hepatitis c
24	138	89.6	74	12 O68679	O68679 hepatitis c
25	138	89.6	74	12 O68683	O68683 hepatitis c
26	138	89.6	74	12 O68706	O68706 hepatitis c
27	138	89.6	74	12 O68684	O68684 hepatitis c
28	138	89.6	74	12 O68707	O68707 hepatitis c
29	138	89.6	74	12 O68685	O68685 hepatitis c
30	138	89.6	76	12 O8JYS0	O8JYS0 hepatitis c
31	138	89.6	78	12 O8JYS1	O8JYS1 hepatitis c
32	138	89.6	83	12 O81264	O81264 hepatitis c
33	138	89.6	100	12 O9QT61	O9QT61 hepatitis c
34	138	89.6	100	12 O8QP70	O8QP70 hepatitis c
35	138	89.6	100	12 O8QP54	O8QP54 hepatitis c
36	138	89.6	100	12 O9QT57	O9QT57 hepatitis c
37	138	89.6	100	12 O8QP52	O8QP52 hepatitis c
38	138	89.6	100	12 O8QP88	O8QP88 hepatitis c
39	138	89.6	100	12 O9QT56	O9QT56 hepatitis c
40	138	89.6	100	12 O8QP90	O8QP90 hepatitis c
41	138	89.6	100	12 O8QP64	O8QP64 hepatitis c
42	138	89.6	100	12 O8QP55	O8QP55 hepatitis c
43	138	89.6	100	12 O8QP60	O8QP60 hepatitis c
44	138	89.6	100	12 O8QP57	O8QP57 hepatitis c
45	138	89.6	100	12 O9QT58	O9QT58 hepatitis c

ALIGNMENTS

RESULT 1

ID P87751 PRELIMINARY; PRT; 415 AA.
AC P87751;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core, env, and part of E2/NS1 (Genome polypeptide) (Fragment).
OS Hepatitis C virus
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VN235;
RA Okamoto H.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VN235;
RX MEDLINE=95062197; PubMed=7972001.
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the
RT seventh, eighth, and ninth major genetic groups.",
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; D88467; BAA13616.1; .
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 415 415
 SQ SEQUENCE 415 AA; 45301 MW; 39B2040A86517D1B CRC64;

Query Match 100.0%; Score 154; DB 12; Length 415;
 Best Local Similarity 100.0%; Pred. No. 1.7e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPQKRNQNTNRRPDVKFPGGGQIVG 28
 Db 5 PKPQKRNQNTNRRPDVKFPGGGQIVG 32

RESULT 2
 O92530
 ID O92530 PRELIMINARY; PRT; 3013 AA.
 AC O92530;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Genome polyprotein.
 DE Hepatitis C virus.
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN235;
 RA Okamoto H.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VN235;
 RX MEDLINE=98378034; PubMed=9714232;
 RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F., Miyakawa Y.,
 RA Yamui M.;
 RT "The entire nucleotide sequences of three hepatitis C virus isolates
 RT in genetic groups 7-9 and comparison with those in the other eight
 RT J. Gen. Virol. 79:1847-1857(1998).
 RL [1]
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; D84263; BAA32665.1; -;
 DR HSP; P27958; IALV.
 DR MEROPS; S29.001; -;
 DR MEROPS; U39.001; -;
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002186; HCV_RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.

DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS05007; RDRP_POSITIVE; 1.
 DR PROSITE; PS05021; RDRP_VIRAL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3013 AA; 328196 MW; C9EE9C0231E86EAF CRC64;

Query Match 100.0%; Score 154; DB 12; Length 3013;
 Best Local Similarity 100.0%; Pred. No. 1.5e-13;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPQKRNQNTNRRPDVKFPGGGQIVG 28
 Db 5 PKPQKRNQNTNRRPDVKFPGGGQIVG 32

RESULT 3
 Q8V7Y2
 ID Q8V7Y2 PRELIMINARY; PRT; 191 AA.
 AC Q8V7Y2;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Core protein (Genome polyprotein) (Fragment).
 DE Hepatitis C virus.
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Kato N.;
 RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous
 RT lesions".
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AB061997; BAB83336.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 KW Polyprotein.
 FT NON_TER 191 191
 SQ SEQUENCE 191 AA; 20703 MW; 3215979FE3F45EB7 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 191;
 Best Local Similarity 92.9%; Pred. No. 1.2e-13;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQKRNQNTNRRPDVKFPGGGQIVG 28
 Db 5 PKPQKRNQNTNRRPDVKFPGGGQIVG 32

RESULT 4
 Q68137
 ID Q68137 PRELIMINARY; PRT; 191 AA.
 AC Q68137;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Genome polyprotein (Fragment).
 DE Hepatitis C virus.
 DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SA6;

```

RX MEDLINE-94336721; PubMed-8058787;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the core gene of 14 hepatitis C virus
RL genotypes.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SA6;
RX MEDLINE-92279243; PubMed-1317578;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the 5' noncoding region of hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SA6;
RX MEDLINE-93376778; PubMed-8396266;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "At least 12 genotypes of hepatitis C virus predicted by sequence
RL analysis of the putative E1 gene of isolates collected worldwide.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
RN [1]
RP SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U10220; AAA21059.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Polyprotein. 191 191
FT NON_TER
SQ SEQUENCE 191 AA; 20682 MW; 0EB8000CF1F902EB CRC64;
Query Match 92.2%; Score 142; DB 12; Length 191;
Best Local Similarity 89.3%; Pred. No. 4.5e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKRNQRTNRRPDVRFPGGGQIVG 28
DB 5 PKPQKRTNRRPDVRFPGGGQIVG 32
||||: |||||||||
RESULT 5
Q9E983 PRELIMINARY; PRT; 103 AA.
AC Q9E983;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
OC Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-N3;
RX MEDLINE-20440673; PubMed-10970401;
RA Tokita H., Kaufmann G.R., Matsubayashi M., Okuda I., Tanaka T.,
RA Harada H., Mukaide M., Suzuki K., Cooper D.A.;
RT "Hepatitis C Virus Core Mutations Reduce the Sensitivity of a
RT Fluorescence Enzyme Immunoassay.";
RL J. Clin. Microbiol. 38:3450-3452(2000).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AB039869; BAB12417.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
DR Polyprotein.
FT NON_TER
SQ SEQUENCE 103 103

```

```

SQ SEQUENCE 103 AA; 11641 MW; B59832ACDA5085EA CRC64;
Query Match 91.6%; Score 141; DB 12; Length 103;
Best Local Similarity 89.3%; Pred. No. 3.3e-13;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKRNQRTNRRPDVRFPGGGQIVG 28
DB 2 PKPQKRTNRRPDVRFPGGGQIVG 29
||||: |||||||||
RESULT 6
Q9J3H8 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OC Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MD14;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RL progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF207755; AAF65945.1; -.
DR HSSP: P26663; LJXP.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00398; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS00507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326421 MW; B109B6487CD20658 CRC64;
Query Match 91.6%; Score 141; DB 12; Length 3010;
Best Local Similarity 89.3%; Pred. No. 1.2e-11;
Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 PKPQKRNQNRNRRPQDVKPPGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKRNTNRRPQDVKPPGGQIVG 32

RESULT 7
 Q90IY9 PRELIMINARY; PRT; 3010 AA.
 ID Q90IY9
 AC Q90IY9
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 CC Hepatitis C virus.
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MD1-2;
 MEDLINE=20013325; PubMed=10544098;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
 Tazawa J.I., Izumi N., Marumo F., Sato C.;
 RT *Time-related changes in full-length hepatitis C virus and hepatitis
 activity.*;
 RL Virology 263:244-253(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MD1-2;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
 Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: AF165046; AAD56181.1; -;
 DR HSP: P26663; LJXP
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS05057; RDRP_POSITIVE; 1.
 DR PROSITE: PS05021; RDRP_VIRAL; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326366 MW; D816D3BBBF14EE46 CRC64;
 Query Match 91.68; Score 141; DB 12; Length 3010;
 Best Local Similarity 89.38; Pred. No. 1.2e-11;
 Matches 25; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNRNRRPQDVKPPGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKRNTNRRPQDVKPPGGQIVG 32

RESULT 8
 Q8V7V3 PRELIMINARY; PRT; 191 AA.
 ID Q8V7V3
 AC Q8V7V3
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Core protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 CC Hepatitis C virus.
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Liver;
 RA Kato N.;
 RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous
 lesions.*";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: AB062026; BAB83365.1; -;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 KW Polyprotein.
 FT NON_TER 191
 SQ SEQUENCE 191 AA; 20700 MW; DDEF95E878228E07 CRC64;
 Query Match 90.98; Score 140; DB 12; Length 191;
 Best Local Similarity 89.38; Pred. No. 8.9e-13;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKRNQNRNRRPQDVKPPGGQIVG 28
 ||||| :|||||
 Db 5 PKPQRTKRNTNRRPQDVKPPGGQIVG 32

RESULT 9
 Q8V7W0 PRELIMINARY; PRT; 191 AA.
 ID Q8V7W0
 AC Q8V7W0
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Core protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 CC Hepatitis C virus.
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Liver;
 RA Kato N.;
 RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous
 lesions.*";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: AB062019; BAB83358.1; -;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR Pfam: PF01543; HCV_capsid; 1.

```
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein. 191 191
FT NON_TER 191 AA; DDEF9785E8228E07 CRC64;
SQ SEQUENCE 191 AA; 20690 MW; 20690 MW; DDEF9785E8228E07 CRC64;

Query Match 90.9%; Score 140; DB 12; Length 191;
Best Local Similarity 89.3%; Pred. No. 8.9e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRNPQDVKPPGGQIVG 28
   |||||: |||||: |||||: |||||: |||||:
Db 5 PKPQKTNRTNRNPQDVKPPGGQIVG 32

RESULT 10
Q8V7W8 PRELIMINARY; PRT; 191 AA.
AC Q8V7W8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous
RT lesions.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AB062011; BAB83350.1; -.
DR InterPro: IPR002522; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein. 191 191
FT NON_TER 191 AA; 20667 MW; 1EEF95EA0E115BFA CRC64;
SQ SEQUENCE 191 AA; 20667 MW; 1EEF95EA0E115BFA CRC64;

Query Match 90.9%; Score 140; DB 12; Length 191;
Best Local Similarity 89.3%; Pred. No. 8.9e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRNPQDVKPPGGQIVG 28
   |||||: |||||: |||||: |||||: |||||:
Db 5 PKPQKTNRTNRNPQDVKPPGGQIVG 32

RESULT 11
Q8V7W5 PRELIMINARY; PRT; 191 AA.
AC Q8V7W5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous
RT lesions.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AB062011; BAB83350.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein. 191 191
FT NON_TER 191 AA; 20667 MW; 1EEF95EA0E115BFA CRC64;
SQ SEQUENCE 191 AA; 20667 MW; 1EEF95EA0E115BFA CRC64;

Query Match 90.9%; Score 140; DB 12; Length 191;
Best Local Similarity 89.3%; Pred. No. 8.9e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRNPQDVKPPGGQIVG 28
   |||||: |||||: |||||: |||||: |||||:
Db 5 PKPQKTNRTNRNPQDVKPPGGQIVG 32

RESULT 12
Q8JWL7 PRELIMINARY; PRT; 191 AA.
AC Q8JWL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L27;
RA Ogata S.;
RT "Comparative Sequence Analysis of the Core Protein and Its Frameshift
RT Product, the F Protein, of Hepatitis C Virus Subtype 1b Strains
RT Obtained from Patients with and without Hepatocellular Carcinoma.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AB077728; BAC02463.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein. 1 1
FT NON_TER 191 191
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20804 MW; F5DCDB21037433C8 CRC64;

Query Match 90.3%; Score 139; DB 12; Length 191;
Best Local Similarity 85.7%; Pred. No. 1.3e-12;
Matches 24; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRNPQDVKPPGGQIVG 28
   |||||: |||||: |||||: |||||: |||||:
Db 5 PKPQKSKRNTNRNPQDVKPPGGQIVG 32

RESULT 13
Q68311 PRELIMINARY; PRT; 45 AA.
AC Q68311;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
```

```
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AB062014; BAB83353.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein. 191 191
FT NON_TER 191 191
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20690 MW; DDEF9785E8228E07 CRC64;

Query Match 90.9%; Score 140; DB 12; Length 191;
Best Local Similarity 89.3%; Pred. No. 8.9e-13;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRNPQDVKPPGGQIVG 28
   |||||: |||||: |||||: |||||: |||||:
Db 5 PKPQKTNRTNRNPQDVKPPGGQIVG 32
```

```
RESULT 12
Q8JWL7 PRELIMINARY; PRT; 191 AA.
AC Q8JWL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L27;
RA Ogata S.;
RT "Comparative Sequence Analysis of the Core Protein and Its Frameshift
RT Product, the F Protein, of Hepatitis C Virus Subtype 1b Strains
RT Obtained from Patients with and without Hepatocellular Carcinoma.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AB077728; BAC02463.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein. 1 1
FT NON_TER 191 191
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20804 MW; F5DCDB21037433C8 CRC64;

Query Match 90.3%; Score 139; DB 12; Length 191;
Best Local Similarity 85.7%; Pred. No. 1.3e-12;
Matches 24; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRNPQDVKPPGGQIVG 28
   |||||: |||||: |||||: |||||: |||||:
Db 5 PKPQKSKRNTNRNPQDVKPPGGQIVG 32

RESULT 13
Q68311 PRELIMINARY; PRT; 45 AA.
AC Q68311;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
```

DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB37;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U23749; AAA65056.1; -;
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5015 MW; CC527167096AA81 CRC64;
Query Match 89.6%; Score 138; DB 12; Length 45;
Best Local Similarity 85.7%; Pred. No. 3.7e-13;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKQKRNQNTNRRPQDVKFPGGQIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32
RESULT 14
O68306 PRELIMINARY; PRT; 45 AA.
AC Q68306;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB5;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U23744; AAA65051.1; -;
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5030 MW; CC4C21ED236AAA81 CRC64;
Query Match 89.6%; Score 138; DB 12; Length 45;
Best Local Similarity 85.7%; Pred. No. 3.7e-13;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKQKRNQNTNRRPQDVKFPGGQIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32
RESULT 15
O68309 PRELIMINARY; PRT; 46 AA.
AC Q68309;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB10;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U23747; AAA65054.1; -;
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 46 AA; 5129 MW; ABC4C21ED236AAA CRC64;
Query Match 89.6%; Score 138; DB 12; Length 46;
Best Local Similarity 85.7%; Pred. No. 3.8e-13;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKQKRNQNTNRRPQDVKFPGGQIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32
Search completed: August 7, 2003, 11:19:01
Job time : 25.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(Without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-29
Perfect score: 154
Sequence: 1 PKPQRNQRNRRPDVKFPGGGIIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued Patents AA:*
- 1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*
 - 2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*
 - 3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*
 - 4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*
 - 5: /cgn2.6/ptodata/1/1aa/PTUS_COMB.pep.*
 - 6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	100.0	28	3	US-08-921-887-29
2	147	95.5	35	1	US-07-681-703B-26
3	147	95.5	35	2	US-08-407-410B-26
4	147	95.5	35	2	US-08-485-500-26
5	147	95.5	35	5	PCT-US91-02370-26
6	147	95.5	75	5	PCT-US92-07813-1
7	147	95.5	90	1	US-07-681-703B-20
8	147	95.5	90	2	US-08-407-410B-20
9	147	95.5	90	2	US-08-485-500-20
10	147	95.5	90	5	PCT-US91-02370-20
11	147	95.5	105	3	US-08-444-818-107
12	147	95.5	119	1	US-07-681-703B-18
13	147	95.5	119	2	US-08-407-410B-18
14	147	95.5	119	2	US-08-485-500-18
15	147	95.5	119	5	PCT-US91-02370-18
16	147	95.5	120	4	US-08-850-328-2
17	147	95.5	144	3	US-08-444-818-103
18	147	95.5	150	1	US-07-681-703B-16
19	147	95.5	150	2	US-08-407-410B-16
20	147	95.5	150	2	US-08-485-500-16
21	147	95.5	150	5	PCT-US91-02370-16
22	147	95.5	215	1	US-07-681-703B-12
23	147	95.5	215	1	US-07-681-703B-14
24	147	95.5	215	5	PCT-US91-02370-12
25	147	95.5	215	5	PCT-US91-02370-14
26	147	95.5	217	2	US-08-407-410B-12
27	147	95.5	217	2	US-08-407-410B-14

28 147 95.5 217 2 US-08-485-500-12 Sequence 12, Appl
29 147 95.5 217 2 US-08-485-500-14 Sequence 14, Appl
30 147 95.5 396 3 US-08-867-611-2 Sequence 7, Appl
31 147 95.5 396 5 PCT-US92-06965A-7 Sequence 12, Appl
32 147 95.5 1021 1 US-07-910-760-12 Sequence 12, Appl
33 147 95.5 1021 1 US-08-440-519-12 Sequence 12, Appl
34 147 95.5 1021 4 US-08-440-549-12 Sequence 12, Appl
35 147 95.5 2995 3 US-08-444-818-133 Sequence 138, App
36 147 95.5 3011 1 US-08-440-103-36 Sequence 36, Appl
37 147 95.5 3011 1 US-08-440-542-36 Sequence 36, Appl
38 147 95.5 3011 1 US-07-910-760-10 Sequence 10, Appl
39 147 95.5 3011 1 US-08-440-519-10 Sequence 10, Appl
40 147 95.5 3011 1 US-08-231-368-36 Sequence 36, Appl
41 147 95.5 3011 1 US-08-440-210-36 Sequence 36, Appl
42 147 95.5 3011 3 US-09-388-874-2 Sequence 2, Appl
43 147 95.5 3011 4 US-09-046-604-36 Sequence 36, Appl
44 147 95.5 3011 4 US-08-440-549-10 Sequence 10, Appl
45 147 95.5 3011 4 US-08-850-328-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-921-887-29
; Sequence 29, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAM
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-29

Query Match 100.0%; Score 154; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 4, 9e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 ZIP: 94300
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: Patent In Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/485,500
16
17 FILING DATE:
18
19 CLASSIFICATION: 435
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: US/08/407,410

```

```

; FILING DATE: US 505,611
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; FILING DATE: 09-OCT-1990
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-500-26

Query Match          95.5%; Score 147; DB 2; Length 35;
Best Local Similarity 92.9%; Pred. No. 5.8e-14;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRRPDVFPFGGGQIVG 28
DB 5 PRPQKNKNTNRRPDVFPFGGGQIVG 32

RESULT 5
PCT-US91-02370-26
; Sequence 26, Application PC/TUS9102370
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; APPLICANT: Simonsen, Christian C.
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: P.O. BOX 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02370
; FILING DATE: 19910405
; CLASSIFICATION: 435.5
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

PCT-US91-02370-26

Query Match          95.5%; Score 147; DB 5; Length 35;
Best Local Similarity 92.9%; Pred. No. 5.8e-14;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRRPDVFPFGGGQIVG 28
DB 5 PRPQKNKNTNRRPDVFPFGGGQIVG 32

RESULT 6
PCT-US92-07813-1
; Sequence 1, Application PC/TUS9207813
; GENERAL INFORMATION:
; APPLICANT: LESNIEWSKI, RICHARD R.
; APPLICANT: LEUNG, TAT K.
; TITLE OF INVENTION: HEPATITIS C ASSAY
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES CHAD377/A36D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.A.
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/07813
; FILING DATE: 19920916
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKIP, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4767.P3.03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US92-07813-1

Query Match          95.5%; Score 147; DB 5; Length 75;
Best Local Similarity 92.9%; Pred. No. 1.3e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKNQNTNRRPDVFPFGGGQIVG 28
DB 5 PRPQKNKNTNRRPDVFPFGGGQIVG 32

RESULT 7
US-07-681-703B-20
; Sequence 20, Application US/07681703B
; Patent No. 5443965
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250

```


CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/07/681,703B
FILING DATE: 05-APR-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 06-APR-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 594,854
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-076.21
TELEPHONE: (415) 324-0880
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-681-703B-20

Query Match 95.5%; Score 147; DB 1; Length 90;
Best Local Similarity 92.9%; Pred. No. 1.5e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PKPKRNQNTNRRPQDVKFGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPKKNKNTNRRPQDVKFGGQIVG 32

RESULT 8
US-08-407-410B-20
Sequence 20, Application US/08407410B
Patent No. 5843636
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory
APPLICANT: Kim, Jungshuh P.
APPLICANT: Moeckli, Randolph
APPLICANT: Simonsen, Christian C.
TITLE OF INVENTION: Hepatitis C Virus Epitopes
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Peter J. Dehlinger
STREET: 350 Cambridge Ave., Suite 100
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/407,410B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 594,854
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-076.21
TELEPHONE: (415) 323-8302
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-407-410B-20

Query Match 95.5%; Score 147; DB 2; Length 90;
Best Local Similarity 92.9%; Pred. No. 1.5e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PKPKRNQNTNRRPQDVKFGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPKKNKNTNRRPQDVKFGGQIVG 32

RESULT 9
US-08-485-500-20
Sequence 20, Application US/084855500
Patent No. 5843639
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory
APPLICANT: Kim, Jungshuh P.
APPLICANT: Moeckli, Randolph
APPLICANT: Simonsen, Christian C.
TITLE OF INVENTION: Hepatitis C Virus Epitopes
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Peter J. Dehlinger
STREET: 350 Cambridge Ave., Suite 100
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/485,500
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/407,410
FILING DATE:
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-076.21
TELEPHONE: (415) 323-8302
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-500-20

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 594,854
;; FILING DATE: 09-OCT-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fabian, Gary R.
;; REGISTRATION NUMBER: 33,875
;; REFERENCE/DOCKET NUMBER: 4600-076.21
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 324-0880
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 119 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-07-681-703B-18

Query Match 95.5%; Score 147; DB 1; Length 119;
Best Local Similarity 92.9%; Pred. No. 2.1e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 1 PKPKRNQNTNRRPDQVKFPGGGQIVG 28
5 PKPKKNKRNTRRRPDQVKFPGGGQIVG 32

RESULT 13
US-08-407-410B-18
; Sequence 18, Application US/08407410B
; Patent No. 5843636
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; APPLICANT: Simonsen, Christian C.
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: 350 Cambridge Ave., Suite 100
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,410B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-407-410B-18

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 505,611
;; FILING DATE: 06-APR-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 594,854
;; FILING DATE: 09-OCT-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fabian, Gary R.
;; REGISTRATION NUMBER: 33,875
;; REFERENCE/DOCKET NUMBER: 4600-076.21
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 323-8302
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 119 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-407-410B-18

Query Match 95.5%; Score 147; DB 2; Length 119;
Best Local Similarity 92.9%; Pred. No. 2.1e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PKPKRNQNTNRRPDQVKFPGGGQIVG 28
Db 5 PKPKKNKRNTRRRPDQVKFPGGGQIVG 32

RESULT 14
US-08-485-500-18
; Sequence 18, Application US/08485500
; Patent No. 5843639
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Moeckli, Randolph
; APPLICANT: Simonsen, Christian C.
; TITLE OF INVENTION: Hepatitis C Virus Epitopes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter J. Dehlinger
; STREET: 350 Cambridge Ave., Suite 100
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,500
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,410
; FILING DATE:
; APPLICATION NUMBER: US 505,611
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 594,854
; FILING DATE: 09-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-076.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 323-8302
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-500-18

Query Match 95.5%; Score 147; DB 2; Length 119;
Best Local Similarity 92.9%; Pred. No. 2.1e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PKPKRNQNTNRRPDQVKFPGGGQIVG 28
Db 5 PKPKKNKRNTRRRPDQVKFPGGGQIVG 32

RESULT 15
PCT-US91-02370-18
; Sequence 18, Application PC/TUS9102370
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory
; APPLICANT: Kim, Jungsuh P.

APPLICANT: Moeckli, Randolph
APPLICANT: Simonsen, Christian C.
TITLE OF INVENTION: Hepatitis C Virus Epitopes
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Peter J. Dehlinger
STREET: P.O. BOX 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02370
FILING DATE: 19910405
CLASSIFICATION: 435.5
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 505,611
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 594,854
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-076.41
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 323-8302
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US91-02370-18

Query Match 95.5%; Score 147; DB 5; Length 119;
Best Local Similarity 92.9%; Pred. No. 2,1e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKQKRRNRRPQDVKFPGGQIVG 28
|||||:|||||
Db 5 PKQKRRNRRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 11:23:52
Job time : 11.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146A-29

Perfect score: 154

Sequence: 1 PKPQKNQNTNRRPQDVKFPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	95.5	43	15	US-10-268-562-2
2	147	95.5	122	15	US-10-098-857B-1
3	147	95.5	190	15	US-10-268-562-1
4	147	95.5	3011	9	US-09-916-359-2
5	138	89.6	44	15	US-10-367-677-1
6	138	89.6	74	10	US-09-851-138-10
7	138	89.6	91	9	US-09-758-308-1
8	138	89.6	97	10	US-09-756-875-8
9	138	89.6	103	10	US-09-921-397-77
10	138	89.6	108	10	US-09-851-138-14
11	138	89.6	113	10	US-09-921-397-78
12	138	89.6	137	10	US-09-851-138-46
13	138	89.6	138	10	US-09-851-138-60
14	138	89.6	166	11	US-09-899-046-152
15	138	89.6	166	11	US-09-878-281-152

Sequence 42, Appl
Sequence 44, Appl
Sequence 42, Appl
Sequence 44, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 76, Appl
Sequence 12, Appl
Sequence 54, Appl
Sequence 18, Appl
Sequence 50, Appl
Sequence 52, Appl
Sequence 54, Appl
Sequence 144, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 23, Appl
Sequence 40, Appl
Sequence 4, Appl
Sequence 9, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 3, Appl

16 138 89.6 169 11 US-09-899-046-42
17 138 89.6 169 11 US-09-899-046-44
18 138 89.6 169 11 US-09-878-281-42
19 138 89.6 169 11 US-09-878-281-44
20 138 89.6 182 10 US-09-929-955-2
21 138 89.6 182 10 US-09-929-955-2
22 138 89.6 182 10 US-10-104-966-2
23 138 89.6 191 11 US-09-194-949-3
24 138 89.6 319 10 US-09-851-138-76
25 138 89.6 319 10 US-09-851-138-12
26 138 89.6 319 10 US-09-851-138-18
27 138 89.6 319 11 US-09-899-046-50
28 138 89.6 319 11 US-09-899-046-52
29 138 89.6 319 11 US-09-899-046-54
30 138 89.6 319 11 US-09-899-046-144
31 138 89.6 319 11 US-09-878-281-50
32 138 89.6 319 11 US-09-878-281-52
33 138 89.6 319 11 US-09-878-281-54
34 138 89.6 319 11 US-09-878-281-144
35 138 89.6 809 10 US-09-973-025-50
36 138 89.6 809 11 US-09-899-303-50
37 138 89.6 809 11 US-09-995-860-50
38 138 89.6 809 11 US-09-941-611-23
39 138 89.6 2884 15 US-10-044-995-23
40 138 89.6 2884 15 US-10-259-275-40
41 138 89.6 3011 9 US-09-742-659-4
42 138 89.6 3011 10 US-09-952-572-9
43 138 89.6 3011 10 US-09-929-955-1
44 138 89.6 3011 10 US-09-747-419-20
45 138 89.6 3011 11 US-09-891-894-3

ALIGNMENTS

RESULT 1
US-10-268-562-2
; Sequence 2, Application US/10268562
; Publication No. US20030108563A1
; GENERAL INFORMATION:
; APPLICANT: Otho-Clinical Diagnostics, Inc.
; TITLE OF INVENTION: Reagents for the simultaneous detection of HCV core antigen
; TITLE OF INVENTION: antibodies
; FILE REFERENCE: CDS0287
; CURRENT APPLICATION NUMBER: US/10/268,562
; PRIOR FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: US 60/347,943
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-268-562-2

Query Match 95.5%; Score 147; DB 15; Length 43;
Best Local Similarity 92.9%; Pred. No. 2.2e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Caps 0;
QY 1 PKPQKNQNTNRRPQDVKFPGGQIVG 28
DB 5 PKPQKNQNTNRRPQDVKFPGGQIVG 32

RESULT 2
US-10-098-857B-1
; Sequence 1, Application US/10098857B
; Publication No. US20030032031A1
; GENERAL INFORMATION:
; APPLICANT: GOEDERT, JAMES J.
; APPLICANT: TODD, JOHN A.
; TITLE OF INVENTION: USE OF SPECIFIC ANTIBODY TITERS TO PREDICT HEPATIC

```
; TITLE OF INVENTION: FAILURE IN PEOPLE INFECTED WITH HEPATITIS C VIRUS
; FILE REFERENCE: 1300-1800.01
; CURRENT APPLICATION NUMBER: US/10/098,857B
; CURRENT FILING DATE: 2002-10-15
; PRIOR FILING DATE: 09/616,823
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/143,851
; PRIOR FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-098-857B-1

Query Match          95.5%; Score 147; DB 15; Length 122;
Best Local Similarity 92.9%; Pred. No. 6.6e-13;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
5 PKPKKKNRNTNRRPQDVKFGGGQIVG 32
|||||:|||||:|||||:|||||:|||||

RESULT 3
US-10-268-562-1
; Sequence 1, Application US/10268562
; Publication No. US20030108563A1
; GENERAL INFORMATION:
; APPLICANT: Otho-Clinical Diagnostics, Inc.
; TITLE OF INVENTION: Reagents for the simultaneous detection of HCV core antigens and
; TITLE OF INVENTION: antibodies
; FILE REFERENCE: CDS0287
; CURRENT APPLICATION NUMBER: US/10/268,562
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: US 60/347,943
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-268-562-1

Query Match          95.5%; Score 147; DB 15; Length 190;
Best Local Similarity 92.9%; Pred. No. 1.1e-12;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
5 PKPKKKNRNTNRRPQDVKFGGGQIVG 32
|||||:|||||:|||||:|||||:|||||

RESULT 4
US-09-916-359-2
; Sequence 2, Application US/09916359
; Patent No. US20020034734A1
; GENERAL INFORMATION:
; APPLICANT: Veronique Barban
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
; TITLE OF INVENTION: TREATING C HEPATITIS
; FILE REFERENCE: PMCF97-03A
; CURRENT APPLICATION NUMBER: US/09/916,359
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 09/388,874
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: 97/02,887
; PRIOR FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
```

```
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Virus
US-09-916-359-2

Query Match          95.5%; Score 147; DB 9; Length 3011;
Best Local Similarity 92.9%; Pred. No. 2e-11;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
5 PKPKKKNRNTNRRPQDVKFGGGQIVG 32
|||||:|||||:|||||:|||||:|||||

RESULT 5
US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10/367,677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; TITLE: of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match          89.6%; Score 138; DB 15; Length 44;
Best Local Similarity 85.7%; Pred. No. 3.8e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
4 PKPKKKNRNTNRRPQDVKFGGGQIVG 31
|||||:|||||:|||||:|||||:|||||

RESULT 6
US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOS
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
```

STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851.138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836.075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-851-138-10

Query Match 89.6%; Score 138; DB 10; Length 74;
Best Local Similarity 85.7%; Pred. No. 6.7e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
||||: ||||||||||||||||
Db 5 PKPKRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 7

US-09-758-308-1
Sequence 1, Application US/09758308
Patent No. US20020090607A1
GENERAL INFORMATION:

APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV

TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS

TITLE OF INVENTION: PROTEINS

FILE REFERENCE: 14114.034902

CURRENT APPLICATION NUMBER: US/09/758.308

PRIOR FILING DATE: 2001-01-10

PRIOR APPLICATION NUMBER: 60/092.339

PRIOR FILING DATE: 1999-07-10

NUMBER OF SEQ ID NOS: 5

SOFTWARE: Patentin version 3.0

SEQ ID NO 1

LENGTH: 91

TYPE: PRT

ORGANISM: Hepatitis C Virus

US-09-758-308-1

Query Match 89.6%; Score 138; DB 9; Length 91;
Best Local Similarity 85.7%; Pred. No. 8.3e-12;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
||||: ||||||||||||||||
Db 5 PKPKRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 8

US-09-756-875-8
Sequence 8, Application US/09756875
Patent No. US2002015090A1
GENERAL INFORMATION:

APPLICANT: PIKE, IAN

TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: Suite 701-E Columbia Square

STREET: 555 13th Street, N. W.

CITY: Washington

STATE: D. C.

COUNTRY: U. S.

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/756.875

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/259,721

FILING DATE: 29-AUG-1994

APPLICATION NUMBER: PCT/GB93/00410

FILING DATE: 26-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: ERNST, BARBARA G.

REGISTRATION NUMBER: 30,377

REFERENCE/DOCKET NUMBER: 1808-157A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)783-6040

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 97 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-09-756-875-8

Query Match 89.6%; Score 138; DB 10; Length 97;

Best Local Similarity 85.7%; Pred. No. 8.9e-12;

Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPKRNORNTNRRPQDVKFGGGQIVG 28
||||: ||||||||||||||||

Db 5 PKPKRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 9

US-09-921-397-77

Sequence 77, Application US/09921397

Patent No. US20020151484A1

GENERAL INFORMATION:

APPLICANT: HYBRIGENICS

TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a

TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and

TITLE OF INVENTION: applications thereof

FILE REFERENCE: B4809A - JAZ

CURRENT APPLICATION NUMBER: US/09/921.397

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: EP 00402225.7

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 156

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 77

LENGTH: 103

TYPE: PRT

ORGANISM: Hepatitis C virus

US-09-921-397-77

```

: FILE REFERENCE: B4809A - JAZ
: CURRENT APPLICATION NUMBER: US/09/921.397
: CURRENT FILING DATE: 2001-08-02
: PRIOR APPLICATION NUMBER: EP 0040225.7
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 156
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 78
: LENGTH: 113
: TYPE: PRT
: ORGANISM: Hepatitis C virus
US-09-921-397-78

Query Match      89.6%; Score 138; DB 10; Length 113;
Best Local Similarity 85.7%; Pred. No. 1e-11;
Matches 24; Conservative 3; Mismatches 1; Indels

QY 1 PKPKRNQRNTRRRPQDVKPPGGQIVG 28
Db 5 PKPQRTKRNTRRRPQDVKPPGGQIVG 32
      |||||: :|||||:|||||:|||||:
      |||||: :|||||:|||||:|||||:

RESULT 12
US-09-851-138-46
: Sequence 46, Application US/09851138
: Publication No. US20020183508A1
: GENERAL INFORMATION:
: APPLICANT: MAERTENS, GEERT
: TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS G
: AGENTS
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE & DURKEE
: STREET: P.O. BOX 4433
: CITY: HOUSTON
: STATE: TEXAS
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Microsoft Word 6.0 / ASCII text output
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/851,138
: FILING DATE: 09-May-2001
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/836,075
: FILING DATE: <Unknown>
: APPLICATION NUMBER: EP 94870166.9
: FILING DATE: 21 Oct 1994
: APPLICATION NUMBER: EP 95870076.7
: FILING DATE: 28 Jun 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: KAMMERER, PATRICIA A.
: REGISTRATION NUMBER: 29,775
: REFERENCE/DOCKET NUMBER: INNS:004
: INFORMATION FOR SEQ ID NO: 46:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 137 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46

Query Match      89.6%; Score 138; DB 10; Length 137;
Best Local Similarity 85.7%; Pred. No. 1.3e-11;
Matches 24; Conservative 3; Mismatches 1; Indels

QY 1 PKPKRNQRNTRRRPQDVKPPGGQIVG 28

```


Db 5 PKPQKRNQNTNRRPDVFPFGGGQIVG 32
|||||: |||||

RESULT 13

US-09-851-138-60

; Sequence 60, Application US/09851138

; Publication No. US20020183508A1

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; STUYVER, LIEVEN

; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC

; AGENTS

; NUMBER OF SEQUENCES: 207

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE & DURKEE

; STREET: P.O. BOX 4433

; CITY: HOUSTON

; STATE: TEXAS

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word 6.0 / ASCII text output

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/851,138

; FILING DATE: 09-May-2001

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/836,075

; FILING DATE: <unknown>

; APPLICATION NUMBER: EP 94870166.9

; FILING DATE: 21 Oct 1994

; APPLICATION NUMBER: EP 95870076.7

; FILING DATE: 28 Jun 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: KAMMERER, PATRICIA A.

; REGISTRATION NUMBER: 29,775

; REFERENCE/DOCKET NUMBER: INNS:004

; INFORMATION FOR SEQ ID NO: 60:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 138 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 60:

; i-09-851-138-60

Query Match 89.6%; Score 138; DB 10; Length 138;
Best Local Similarity 85.7%; Pred. No. 1.3e-11;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVFPFGGGQIVG 28

Db 5 PKPQKRNQNTNRRPDVFPFGGGQIVG 32

RESULT 14

US-09-899-046-152

; Sequence 152, Application US/09899046

; Publication No. US20030008274A1

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: New sequences of hepatitis C virus

; genotypes for diagnosis, prophylaxis and therapy.

; NUMBER OF SEQUENCES: 270

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/362,455

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 152:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 166 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-899-046-152

Query Match 89.6%; Score 138; DB 11; Length 166;
Best Local Similarity 85.7%; Pred. No. 1.6e-11;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVFPFGGGQIVG 28

Db 5 PKPQKRNQNTNRRPDVFPFGGGQIVG 32

RESULT 15

US-09-878-281-152

; Sequence 152, Application US/09878281

; Publication No. US20030032005A1

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: New sequences of hepatitis C virus

; genotypes for diagnosis, prophylaxis and therapy.

; NUMBER OF SEQUENCES: 270

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/878,281

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/362,455

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 152:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 166 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-878-281-152

Query Match 89.6%; Score 138; DB 11; Length 166;
Best Local Similarity 85.7%; Pred. No. 1.6e-11;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKRNQNTNRRPDVFPFGGGQIVG 28

Db 5 PKPQKRNQNTNRRPDVFPFGGGQIVG 32

Search completed: August 7, 2003, 12:01:13

Job time : 14.3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds
(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146A-28

Perfect score: 154

Sequence: 1 PKPQRKTRNAHRPDVKFKGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

tal number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	100.0	28	20	AAV06678
2	148	96.1	28	20	AAV06683
3	147	95.5	312	15	AAR58593
4	144	93.5	28	20	AAV06675
5	144	93.5	3010	15	AAR53417
6	144	93.5	3010	23	AAE20477
7	143	92.9	36	16	AAW06487
8	143	92.9	38	14	AAR30687
9	143	92.9	38	14	AAR30688

10	143	92.9	38	14	AAR30689	HCV capsid peptide
11	143	92.9	38	15	AAR54065	Non-A, non-B hepat
12	143	92.9	43	19	AAW37380	Hepatitis C virus
13	143	92.9	44	19	AAW66083	Hepatitis C virus
14	143	92.9	44	20	AAV26952	Hepatitis C virus
15	143	92.9	44	21	AAV94410	Human hepatitis C
16	143	92.9	45	21	AAV94409	Human hepatitis C
17	143	92.9	55	13	AAR29534	HCV core-envelope
18	143	92.9	55	13	AAR29535	HCV core-envelope
19	143	92.9	55	13	AAR29536	HCV core-envelope
20	143	92.9	55	13	AAR29537	HCV core-envelope
21	143	92.9	55	13	AAR29542	HCV core-envelope
22	143	92.9	57	13	AAR25121	Non-A, Non-B Hepat
23	143	92.9	61	13	AAR20770	Peptide VIIIE base
24	143	92.9	61	16	AAR69345	Anti-HCV antibody
25	143	92.9	61	17	AAV15425	Prototype peptide
26	143	92.9	61	18	AAW01865	HCV core protein p
27	143	92.9	66	12	AAR12597	PT-NANB viral stru
28	143	92.9	74	17	AAR96530	Hepatitis C virus
29	143	92.9	78	13	AAR29160	Antigen pHC101.
30	143	92.9	78	13	AAR29161	Antigen pHC101.
31	143	92.9	79	14	AAR41422	HCV fragment 1 / I
32	143	92.9	79	14	AAR41423	HCV fragment 2 / I
33	143	92.9	80	15	AAR51389	Branched peptide H
34	143	92.9	90	16	AAR66619	Hepatitis C virus
35	143	92.9	91	22	AAB31695	Antigenic epitope
36	143	92.9	97	14	AAR40978	HCV core protein N
37	143	92.9	97	16	AAR66633	Hepatitis C virus
38	143	92.9	97	20	AAV01624	Protein encoded by
39	143	92.9	97	20	AAW30597	Hepatitis C virus
40	143	92.9	103	23	ABW77253	HCV bait polypepti
41	143	92.9	105	21	AAB18537	Protein encoded by
42	143	92.9	108	17	AAR65332	Hepatitis C virus
43	143	92.9	113	23	ABW77254	HCV bait polypepti
44	143	92.9	114	17	AAR98348	HCV capsid core pr
45	143	92.9	115	13	AAR29530	HCV core-envelope

ALIGNMENTS

RESULT 1

AAV06678
ID AAV06678 standard; Protein; 28 AA.

AC AAV06678;

DT 17-JUN-1999 (first entry)

DE NC mosaic protein amino acid fragment F.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
restriction endonuclease assisted ligation; vaccination.

OS Hepatitis C virus.

PN WO9910506-A1.

PD 04-MAR-1999.

PF 21-AUG-1998; 98WO-US17385.

PR 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudakov YE;

XX WPI; 1999-204671/17.

PT New mosaic protein, comprising a plurality of homologous antigenic
peptides from different genotypes of a species - useful for
detecting hepatitis infection in an individual

XX

PS Claim 5; Fig 9; 66pp; English.

XX The invention relates to a mosaic protein, comprising a plurality of
 CC homologous antigenic peptides from different genotypes of a species. The
 CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
 CC synthesizing an artificial gene that encodes the mosaic protein is also
 CC provided. The method is designated restriction endonuclease assisted
 CC ligation (REAL). The mosaic protein and the artificial mosaic protein
 CC are useful for detecting a hepatitis infection in an individual. The
 CC mosaic gene and protein is also useful for vaccination against
 CC infection, especially hepatitis C. The method of synthesizing the
 CC artificial gene and the resulting mosaic protein improve the sensitivity,
 CC spectrum of immunoreactivity, and antigen specificity of enzyme
 CC immunoassays. This provides improved detection of hepatitis C virus.
 CC Sequences AAY06673-683 represent amino acid sequence of each monomer
 CC comprising the NC mosaic protein.

XX Sequence 28 AA;

Query Match 100.0%; Score 154; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 2.3e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28

Db 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28

RESULT 2

AAY06683

ID AAY06683 standard; Protein; 28 AA.

XX AC AAY06683;

XX 17-JUN-1999 (first entry)

XX NC mosaic protein amino acid fragment K.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
 KW restriction endonuclease assisted ligation; vaccination.

XX Hepatitis C virus.

XX WO9910506-A1.

XX 04-MAR-1999.

XX 21-AUG-1998; 98WO-US17385.

XX 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1999-204671/17.

XX New mosaic protein, comprising a plurality of homologous antigenic
 PT peptides from different genotypes of a species - useful for
 PT detecting hepatitis infection in an individual

XX Claim 5; Fig 9; 66pp; English.

XX The invention relates to a mosaic protein, comprising a plurality of
 CC homologous antigenic peptides from different genotypes of a species. The
 CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
 CC synthesizing an artificial gene that encodes the mosaic protein is also
 CC provided. The method is designated restriction endonuclease assisted
 CC ligation (REAL). The mosaic protein and the artificial mosaic protein
 CC are useful for detecting a hepatitis infection in an individual. The
 CC mosaic gene and protein is also useful for vaccination against
 CC infection, especially hepatitis C. The method of synthesizing the

CC artificial gene and the resulting mosaic protein improve the sensitivity,
 CC spectrum of immunoreactivity, and antigen specificity of enzyme
 CC immunoassays. This provides improved detection of hepatitis C virus.
 CC Sequences AAY06673-683 represent amino acid sequence of each monomer
 CC comprising the NC mosaic protein.

XX Sequence 28 AA;

Query Match 96.1%; Score 148; DB 20; Length 28;

Best Local Similarity 96.4%; Pred. No. 1.6e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28

Db 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28

RESULT 3

AAR58593

ID AAR58593 standard; Protein; 312 AA.

XX AC AAR58593;

XX 09-MAY-1995 (first entry)

XX Hepatitis C virus 349-1285 fragment antigen.

XX Non-A non-B hepatitis virus antigens; NANBH; hepatitis C virus.

XX Hepatitis C virus.

XX JP06225770-A.

XX 16-AUG-1994.

XX 08-JUL-1993; 93JP-0193104.

XX 10-JUL-1992; 92JP-0207391.

XX (KOKU-) KOKUSAI SHIYAKU KK.

XX (SANW) SANWA KAGAKU KENKYUSHO CO.

XX (TOFU) TONEN CORP.

XX (TOKR-) 2H TOKYO TO RINSHO IGAKU SOGO KENKYUSHO.

XX WPI; 1994-298800/37.

XX N-PSDB; AAQ70543.

XX A nucleic acid fragment coding Non-A Non-B Hepatitis virus
 PT antigens - for diagnosis of NANBH and detection of HCV

XX Claim 1; Page 18; 22pp; Japanese.

XX AAQ70543 is a fragment of hepatitis C virus (HCV) or non-A non-B
 CC hepatitis virus (NANBH) core and envelope one structural regions,
 CC encompassing base pairs 349-1285. It codes for AAR58593 an antigen
 CC to a structural region of the HCV virus, which can be used in
 CC the diagnosis of NANBH patients and the detection of HCV carriers.

XX Sequence 312 AA;

Query Match 95.5%; Score 147; DB 15; Length 312;

Best Local Similarity 96.4%; Pred. No. 2.6e-12;

Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28

Db 2 PKPQKTKRNAHRRPQDVKFPGGQIVG 29

RESULT 4

AAY06675

ID AAY06675 standard; Protein; 28 AA.

XX

```

AC AAY06675;
XX
XX 17-JUN-1999 (first entry)
XX
XX NC mosaic protein amino acid fragment C.
XX
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
XX restriction endonuclease assisted ligation; vaccination.
XX
XX Hepatitis C virus.
XX
XX WO9910506-A1.
XX
XX 04-MAR-1999.
XX
XX 21-AUG-1998; 98WO-US17385.
XX
XX 25-AUG-1997; 97US-0921887.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Fields HA, Khudyakov YE;
XX
XX WPI: 1999-204671/17.
XX
XX New mosaic protein, comprising a plurality of homologous antigenic
XX peptides from different genotypes of a species - useful for
XX detecting hepatitis infection in an individual
XX
XX Claim 5; Fig 9; 66pp; English.
XX
XX The invention relates to a mosaic protein, comprising a plurality of
XX homologous antigenic peptides from different genotypes of a species. The
XX antigenic peptides are from nucleocapsid (NC) proteins. A method for
XX synthesizing an artificial gene that encodes the mosaic protein is also
XX provided. The method is designated restriction endonuclease assisted
XX ligation (REAL). The mosaic protein and the artificial mosaic protein
XX are useful for detecting a hepatitis infection in an individual. The
XX mosaic gene and protein is also useful for vaccination against
XX infection, especially hepatitis C. The method of synthesizing the
XX artificial gene and the resulting mosaic protein improve the sensitivity,
XX spectrum of immunoreactivity, and antigen specificity of enzyme
XX immunoassays. This provides improved detection of hepatitis C virus.
XX Sequences AAY06673-683 represent amino acid sequence of each monomer
XX comprising the NC mosaic protein.
XX
XX Sequence 28 AA;
XX
XX Query Match 93.5%; Score 144; DB 20; Length 28;
XX best Local Similarity 92.9%; Pred. No. 5.6e-13;
XX Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 PKPQRTKRNHRRPDQVFPFGGQIVG 28
XX ||||| :|||||
XX Db 1 PKPQRTKRTYRRPDQVFPFGGQIVG 28
XX
XX RESULT 5
XX AAR53417
XX ID AAR53417 standard; Protein; 3010 AA.
XX
XX AC AAR53417;
XX
XX 17-JAN-1995 (first entry)
XX
XX Blood transmissible NANBHV protein.
XX
XX Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
XX NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
XX C100 antibody; HCV RNA; NS5 region.
XX
XX Non-A, non-B hepatitis virus.
XX

```

```

FH Key Location/Qualifiers
FH Misc-difference 222 /label= His, Arg
FH Misc-difference 226 /label= Cys, Arg
FH Misc-difference 246 /label= Leu, Phe
FH Misc-difference 263 /label= Asp, Asn
FH Misc-difference 291 /label= Phe, Ser
FH Misc-difference 311 /label= Gly, Asp
FH Misc-difference 398 /label= Ser, Arg, Gly
FH Misc-difference 400 /label= Thr, Ala
FH Misc-difference 405 /label= Gln, Pro, Leu
FH Misc-difference 410 /label= Lys, Arg
FH Misc-difference 418 /label= Gly, Asp
FH Misc-difference 430 /label= Asn, Asp
FH Misc-difference 438 /label= Phe, Leu
FH Misc-difference 478 /label= Arg, Lys
FH Misc-difference 759 /label= Leu, Val
FH Misc-difference 1017 /label= Ser, Asn
FH Misc-difference 1036 /label= Thr, Ala
FH Misc-difference 1056 /label= Glu, Asp
FH Misc-difference 1201 /label= Met, Thr
FH Misc-difference 1205 /label= Met, Ile
FH Misc-difference 1255 /label= Asn, Tyr
FH Misc-difference 1263 /label= Gly, Asp
FH Misc-difference 1455 /label= Asn, Asp
FH Misc-difference 1828 /label= Ala, Thr
FH Misc-difference 1895 /label= Gly, Arg
FH Misc-difference 1896 /label= Gly, Ile
FH Misc-difference 2143 /label= Glu, Val
FH Misc-difference 2144 /label= Asp, Glu
FH Misc-difference 2462 /label= Cys, Arg
FH Misc-difference 2486 /label= Val, Met
FH Misc-difference 2488 /label= Lys, Gln
FH Misc-difference 2844 /label= Leu, Met
FH Misc-difference 2862 /label= Leu, Gln
FH Misc-difference 2917 /label= Arg, Leu
FH Misc-difference 2968 /label= Ser, Gly
FH Misc-difference 2989 /label= Cys, Arg
FH

```

FT Misc-difference 2990
 FT /label= Tyr, Cys
 PN JP06105690-A.
 XX 19-APR-1994.
 PD
 XX 10-MAR-1992; 92JP-0051885.
 PF
 XX 10-MAR-1992; 92JP-0051885.
 PR
 XX (KAEN/) KAENNO K.
 PA
 XX WPI; 1994-163130/20.
 XX N-PSDB; AAQ63499.
 DR
 XX Blood-transmissible non-A non-B hepatitis virus DNA - used for
 PT detection of hepatitis virus
 XX
 XX Claim 1; Page 8-20; 22pp; Japanese.
 --
 CC This sequence is encoded by the genome of a blood transmissible non-A,
 CC non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the
 CC primers given in AAQ63500-35. The amplified fragments are used in the
 CC detection of hepatitis virus. The target DNA was isolated from serum
 CC of chronically infected NANBH patients who were C100 antibody-positive
 CC and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR
 CC were performed on cDNA and the total human NANBH DNA was constructed
 CC from 23 clones.
 XX
 XX SQ Sequence 3010 AA;
 Query Match 93.5%; Score 144; DB 15; Length 3010;
 Best Local Similarity 92.9%; Pred. No. 6.9e-11;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28
 DB 5 PKPQKTKRNTYRRPQDVKFPGGQIVG 32
 RESULT 6
 AAEE20477
 ID AAEE20477 standard; Protein; 3010 AA.
 XX
 AC AAEE20477;
 XX
 DT 01-JUL-2002 (first entry)
 HCV-S1 full-length polyprotein.
 ..A
 KW Nucleic acid construct; expression cassette; non-coding region; NCR;
 KW untranslated region; UTR; anti-viral drug; drug resistance;
 KW HCV-S1; Hepatitis C virus.
 XX
 OS Hepatitis C virus.
 XX
 PN WO200208447-A2.
 XX
 PD 31-JAN-2002.
 XX
 XX 20-JUL-2001; 2001WO-IL00669.
 PF
 XX 24-JUL-2000; 2000US-220248P.
 PR
 XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.
 PA (EHLR/) EHLRICH G.
 XX
 XX Tan YH, Lim SP, Lim SG, Hong WJ;
 PI WPI: 2002-280605/32.
 XX N-PSDB; AAD33038.
 DR
 XX

PT Novel nucleic acid construct useful for detecting the presence of RNA
 PT virus, comprises an expression cassette and a promoter operably linked
 PT to expression cassette for minus strand RNA transcription of the
 PT cassette
 XX
 PS Example 1; Page 70-81; 81pp; English.
 XX
 CC The invention relates to nucleic acid construct which comprises an
 CC expression cassette including a first polynucleotide region including
 CC a 5' non-coding region (NCR) sequence of an RNA virus and at least an
 CC N-terminal portion of a coding sequence of RNA virus, a second
 CC polynucleotide region including a 3' untranslated region (UTR) sequence
 CC of the RNA virus and at least a C-terminal portion of a coding sequence
 CC of the RNA virus, and a third polynucleotide region encoding a reporter
 CC molecule, flanked by first and second polynucleotide regions; and a
 CC promoter sequence being operatively linked to expression cassette in a
 CC manner so as to enable a transcription of a minus strand RNA molecule
 CC from the expression cassette. Nucleic acid construct of the invention
 CC is useful for detecting the presence of an RNA virus in a cell. It is
 CC also useful for screening anti-viral drugs and determining drug
 CC resistance of an RNA virus. The present sequence is Hepatitis C virus
 CC (HCV) isolate HCV-S1 full-length polyprotein.
 XX
 XX SQ Sequence 3010 AA;
 Query Match 93.5%; Score 144; DB 23; Length 3010;
 Best Local Similarity 92.9%; Pred. No. 6.9e-11;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28
 DB 5 PKPQKTKRNTYRRPQDVKFPGGQIVG 32
 RESULT 7
 AAWE06487
 ID AAWE06487 standard; Protein; 35 AA.
 XX
 AC AAWE06487;
 XX
 DT 31-JAN-1997 (first entry)
 XX
 DE CN14 fragment of HCV core region.
 XX
 XX CN14; CP14; core region; hepatitis C virus; HCV; detection;
 KW antibody.
 XX
 OS Synthetic.
 XX
 PN JP06327482-A.
 XX
 PD 29-NOV-1994.
 XX
 XX 21-MAY-1993; 93JP-0156026.
 PF
 XX 21-MAY-1993; 93JP-0156026.
 PR
 XX (IMMO) IMMUNO JAPAN KK.
 XX
 XX WPI; 1995-047903/07.
 DR N-PSDB; AAT45055.
 DR
 XX Detection of hepatitis C virus - using oligopeptide fragment of HCV
 PT core region
 XX
 PS Claim 1; Page 6; 7pp; Japanese.
 XX
 CC This sequence is encoded by the oligonucleotide, CN14, and represents
 CC the peptide fragment CP14. CP14 is a fragment of the core region
 CC of hepatitis C virus (HCV). CP14 may be used in the detection of
 CC HCV infection and to raise antibodies against it.
 XX
 XX SQ Sequence 36 AA;

```
Query Match          92.9%; Score 143; DB 16; Length 36;
Best Local Similarity 92.9%; Pred. No. 1e-12; 1; Indels 0; Gaps 0;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFFGGGQIVG 28
    ||||| :|||||
Db 1 PKPQKTKRNTNRRPQDVKFFGGGQIVG 28
    ||||| :|||||

RESULT 8
AAR30687
ID AAR30687 standard; peptide; 38 AA.
XX
AC AAR30687;
XX
XX 25-MAR-2003 (updated)
DT 11-MAY-1993 (first entry)
XX
XX HCV capsid peptide No. 23.
XX
KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;
KW non-A, non-B hepatitis; competitive; inhibition assay.
XX
OS Hepatitis C virus.
XX
PN W09222571-A1.
XX
XX 23-DEC-1992.
XX
XX 29-APR-1992; 92WO-US03635.
XX
PR 13-JUN-1991; 91US-0714471.
PR 20-JUN-1991; 91US-0718052.
XX
XX (BAXT ) BAXTER DIAGNOSTICS INC.
XX
PI Jolley ME, Leahy DC, Todd JA;
XX
XX WPI; 1993-018073/02.
XX
XX Synthetic peptide epitope with sequence encoded by hepatitis-C
XX virus - for immunoassay for antigens for diagnosis of non-A,
XX non-B hepatitis
XX
PS Disclosure; Fig 1F; 66pp; English.
XX
CC The sequences given in AAR30665-89 represent fragments of the
CC Hepatitis C virus (HCV) amino acid sequence. They represent the
CC beginning of the HCV open reading frame to amino acid 38 and
CC encompass the "common" sequence. These peptides are contained in
CC the capsid protein of the virus and themselves contain epitope
CC groups. These peptides can be used in immunoassays for HCV
CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
CC in competitive inhibition assay for detecting HCV specific
CC antibodies.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 38 AA;

-- Query Match          92.9%; Score 143; DB 14; Length 38;
-- Best Local Similarity 92.9%; Pred. No. 1.1e-12;
-- Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFFGGGQIVG 28
    ||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32
    ||||| :|||||

RESULT 9
AAR30688
ID AAR30688 standard; peptide; 38 AA.
XX
XX
XX 25-MAR-2003 (updated)
DT 11-MAY-1993 (first entry)
XX
XX HCV capsid peptide No. 25.
XX
KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;
KW non-A, non-B hepatitis; competitive; inhibition assay.
XX
OS Hepatitis C virus.
```

```
XX AAR30688;
AC
XX 25-MAR-2003 (updated)
DT 11-MAY-1993 (first entry)
XX
XX HCV capsid peptide No. 24.
XX
KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;
KW non-A, non-B hepatitis; competitive; inhibition assay.
XX
OS Hepatitis C virus.
XX
PN W09222571-A1.
XX
XX 23-DEC-1992.
XX
XX 29-APR-1992; 92WO-US03635.
XX
PR 13-JUN-1991; 91US-0714471.
PR 20-JUN-1991; 91US-0718052.
XX
XX (BAXT ) BAXTER DIAGNOSTICS INC.
XX
PI Jolley ME, Leahy DC, Todd JA;
XX
XX WPI; 1993-018073/02.
XX
XX Synthetic peptide epitope with sequence encoded by hepatitis-C
XX virus - for immunoassay for antigens for diagnosis of non-A,
XX non-B hepatitis
XX
PS Disclosure; Fig 1F; 66pp; English.
XX
CC The sequences given in AAR30665-89 represent fragments of the
CC Hepatitis C virus (HCV) amino acid sequence. They represent the
CC beginning of the HCV open reading frame to amino acid 38 and
CC encompass the "common" sequence. These peptides are contained in
CC the capsid protein of the virus and themselves contain epitope
CC groups. These peptides can be used in immunoassays for HCV
CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
CC in competitive inhibition assay for detecting HCV specific
CC antibodies.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 38 AA;

Query Match          92.9%; Score 143; DB 14; Length 38;
Best Local Similarity 92.9%; Pred. No. 1.1e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFFGGGQIVG 28
    ||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFFGGGQIVG 32
    ||||| :|||||

RESULT 10
AAR30689
ID AAR30689 standard; peptide; 38 AA.
XX
XX
XX 25-MAR-2003 (updated)
DT 11-MAY-1993 (first entry)
XX
XX HCV capsid peptide No. 25.
XX
KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;
KW non-A, non-B hepatitis; competitive; inhibition assay.
XX
OS Hepatitis C virus.
```

```

XX PN WO2222571-A1.
XX PD 23-DEC-1992.
XX PF 29-APR-1992; 92WO-US03635.
XX PR 13-JUN-1991; 91US-0714471.
XX PR 20-JUN-1991; 91US-0718052.
XX PA (BAXT ) BAXTER DIAGNOSTICS INC.
XX PI Jolley ME, Leahy DC, Todd JA;
XX PR WPI: 1993-018073/02.
XX PT Synthetic peptide epitope with sequence encoded by hepatitis-C
XX PT virus - for immunoassay for antigens for diagnosis of non-A,
XX PT non-B hepatitis
XX PT Disclosure; Fig 1F; 66pp; English.
XX CC The sequences given in AAR30665-89 represent fragments of the
XX CC Hepatitis C virus (HCV) amino acid sequence. They represent the
XX CC beginning of the HCV open reading frame to amino acid 38 and
XX CC encompass the "common" sequence. These peptides are contained in
XX CC the capsid protein of the virus and themselves contain epitope
XX CC groups. These peptides can be used in immunoassays for HCV
XX CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
XX CC in competitive inhibition assay for detecting HCV specific
XX CC antibodies.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 38 AA;
XX Query Match 92.9%; Score 143; DB 14; Length 38;
XX Best Local Similarity 92.9%; Pred. No. 1.1e-12;
XX Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 PKPQRTKRNAHRRPDVKFPGGGQIVG 28
XX DB ||||| :|||||
XX 5 PKPQRTKRNTRRRPDVKFPGGGQIVG 32
XX RESULT 11
XX AAR54065
XX ID AAR54065 standard; Protein; 38 AA.
XX AAR54065;
XX 14-FEB-1995 (first entry)
XX DE Non-A, non-B hepatitis virus coreI region fragment.
XX KW Non-A, non-B hepatitis virus; NANBH; hepatitis C virus; HCV;
XX KW core; ENV; NS1; NS2; NS3; antigen; detection.
XX XX Hepatitis C virus.
XX OS
XX PN JP06141870-A.
XX XX 24-MAY-1994.
XX PD
XX PF 12-MAR-1992; 92JP-0088140.
XX PR 12-MAR-1992; 92JP-0088140.
XX PA (SANW ) SANWA KAGAKU KENKYUSHO CO.
XX PA (TOFU ) TONEN CORP.
XX PA (TOKR-) 2H TOKYO RINSHO IGAKU SOGO KENKYUSHO.
XX WPI: 1994-205028/25.
XX DR N-PSDB; AAQ64067.
XX DR

```

```

XX XX DNA coding a Non-A, non-B hepatitis virus antigen - useful for
XX PT detecting HCV within serum
XX PS Disclosure; Page 11; 22pp; Japanese.
XX XX Hepatitis C virus #4 and #6 genes were isolated (AAQ64068-69).
XX CC Both genes contain the core, ENV, NS1, NS2 and NS3 regions.
XX CC A core region fragment is given in AAQ64067.
XX XX Sequence 38 AA;
XX Query Match 92.9%; Score 143; DB 15; Length 38;
XX Best Local Similarity 92.9%; Pred. No. 1.1e-12;
XX Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 PKPQRTKRNAHRRPDVKFPGGGQIVG 28
XX DB ||||| :|||||
XX 5 PKPQRTKRNTRRRPDVKFPGGGQIVG 32
XX RESULT 12
XX AAW37380
XX ID AAW37380 standard; Protein; 43 AA.
XX XX AAW37380;
XX AC
XX DT 11-MAR-1998 (first entry)
XX XX Hepatitis C virus C-1 protein 1-43.
XX DE
XX XX Hepatitis C virus; HCV; chimeric; antigen; detection; core region;
XX KW epitope; NS3; NS4; infection.
XX XX Hepatitis c virus.
XX OS
XX PN JP09278794-A.
XX XX 28-OCT-1997.
XX PF 10-FEB-1997; 97JP-0027015.
XX PR 09-FEB-1996; 96JP-0024045.
XX PA (TOFU ) TONEN CORP.
XX XX WPI: 1998-022248/03.
XX XX New chimeric peptide antigen derived from hepatitis C virus protein
XX PT - useful for detecting HCV infections
XX PS Disclosure; Page 24; 30pp; Japanese.
XX XX The present sequence represents a Hepatitis C virus (HCV) protein
XX CC sequence from the disclosure of the present specification. The
XX CC present specification describes a chimeric HCV peptide antigen which
XX CC comprises at least 2 peptide epitope regions from the HCV polyptide
XX CC core region, 2 peptide epitope regions from the NS3 region and at
XX CC least 2 peptide epitope regions from the NS4 region. The antigen binds
XX CC specifically with an antibody produced by a human infected by HCV. The
XX CC peptide can detect a wide range of HCV infections with high sensitivity.
XX XX Sequence 43 AA;
XX Query Match 92.9%; Score 143; DB 19; Length 43;
XX Best Local Similarity 92.9%; Pred. No. 1.2e-12;
XX Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 PKPQRTKRNAHRRPDVKFPGGGQIVG 28
XX DB ||||| :|||||
XX 5 PKPQRTKRNTRRRPDVKFPGGGQIVG 32

```

RESULT 13

AAW66083
ID AAW66083 standard; peptide; 44 AA.

XX AC AAW66083;
XX DT 16-NOV-1998 (first entry)

XX DE Hepatitis C virus p21 protein N-terminal fragment.
XX KW antigenic; nucleocapsid; p21 protein; alpha-helical; immunogen;
KW antibody; hepatitis C virus; epitope; N-terminal fragment.
XX OS hepatitis c virus.
XX PN WO9839360-A1.
XX PD 11-SEP-1998.
XX PF 05-MAR-1998; 98WO-FR00442.
XX PF 05-MAR-1997; 97FR-0002878.
(INMR) BIO MERIEUX.

AX PA Dalbon P, Jolivet M, Lacoux X, Ladaviere L, Penin F;
XX PI WPI; 1998-495793/42.
XX DR New peptide from the N-terminus of hepatitis C virus p21 protein
PT containing the immunodominant epitope - and related antibodies, used
PT for diagnosis, treatment and prevention of hepatitis C infection
XX PS Disclosure; Page 16; 37pp; French.

XX The invention relates to a peptide, which is recognised by antibodies
CC against amino acids 2-45 at the N-terminus of the core (or nucleocapsid)
CC p21 protein of hepatitis C virus (HCV), or its variants. The peptide has
CC a tertiary structure consisting of two alpha-helical fragments, almost
CC perpendicular to each other in space, connected by a junction peptide.
CC Excluded are all proteins and peptides comprising, or consisting of, the
CC N-terminal part of p21 (starting from amino acid 1 or 2). Also new are
CC (1) monoclonal or polyclonal antibodies produced using the peptide as
CC an immunogen and (2) complex consisting of the peptide specifically
CC bound to some other molecule, particularly peptide or nucleotide
CC fragments or functionalised aromatic compounds. The peptide is used (i)
CC as immunogen for generating antibodies and (ii) for detecting and
CC quantifying either antibodies against p21 or HCV-derived mRNA (by
CC complex formation). Antibodies are used correspondingly to detect HCV or
CC related antigens. The peptide and antibodies may also be used to treat
CC or prevent HCV infections. The present sequence represents the N-terminus
XX of the core p21 protein of hepatitis C virus.
SQ Sequence 44 AA;

Query Match 92.9%; Score 143; DB 19; Length 44;
Best Local Similarity 92.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRKRNHRPDQVKFPGGGQIVG 28
DB 4 PKPQKTRKRNHRPDQVKFPGGGQIVG 31

RESULT 14

AAV26952
ID AAV26952 standard; peptide; 44 AA.

XX AC AAV26952;
XX DT 21-DEC-1999 (first entry)

XX DE Hepatitis C virus Core protein amino acids 2-45.

XX

KW Epitope; hepatitis C virus; core protein; monoclonal antibody;
KW diagnosis; infection; sandwich immunoassay.

XX OS Hepatitis C virus.
XX PN FR2775690-A1.
XX PD 10-SEP-1999.
XX PF 09-MAR-1998; 98FR-0003087.
XX PR 09-MAR-1998; 98FR-0003087.
XX PA (INMR) BIO MERIEUX.
XX PI Jolivet RC, Piga N, Yvon S, Paranhos BC, Jolivet M;
XX WPI; 1999-530397/45.
XX DR Monoclonal antibodies useful for detecting and/or quantifying hepatitis
PT C virus core protein
XX PS Claim 3; Page 10; 19pp; French.

XX Peptides AAY26949-Y26955 represent peptide epitopes derived from the
CC N-terminus of the hepatitis C virus core protein. The peptides are used
CC to generate monoclonal antibodies or antibody fragments specific for
CC hepatitis C virus (HCV) core protein. The monoclonal antibodies are
CC used for early diagnosis of HCV infections, especially by two-antibody
CC sandwich immunoassay.
XX SQ Sequence 44 AA;

Query Match 92.9%; Score 143; DB 20; Length 44;
Best Local Similarity 92.9%; Pred. No. 1.2e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRKRNHRPDQVKFPGGGQIVG 28
DB 4 PKPQKTRKRNHRPDQVKFPGGGQIVG 31

RESULT 15

AAV94410
ID AAV94410 standard; peptide; 44 AA.

XX AC AAV94410;
XX DT 11-SEP-2000 (first entry)

XX DE Human hepatitis C virus core protein N-terminus, residues 2-45.
XX KW Human; hepatitis C virus; HCV; immunogen; antigen; antibody; virucide;
KW hepatotropic; anti-inflammatory; virus detection; vaccine.
XX OS Hepatitis C virus.
XX PN WO200031130-A1.
XX PD 02-JUN-2000.
XX PF 19-NOV-1999; 99WO-IB01933.
XX PR 20-NOV-1998; 98US-0196155.
XX PA (INMR) BIO MERIEUX.
XX PI Dalbon P, I Dalbon P, Jolivet M, Jolivet-Reynaud C;
XX WPI; 2000-411934/35.
XX PT Polypeptides that bind to anti-hepatitis C virus antibodies, useful for

PT diagnosing and preventing hepatitis C infections -

XX PS Claim 1; Page 42; 50pp; English.

XX CC The present peptide, designated S42G, corresponds to residues 2 to 45 of
CC the N-terminus of the human hepatitis C virus (HCV) core protein. It is
CC an immunodominant region containing conformational type epitopes and
CC linear type epitopes. It manifests an immunoreactivity with all the sera
CC of individuals or blood samples infected with HCV and which possess
CC antibodies directed against the core protein. An amino acid may be
CC substituted for homologous amino acids and side chains and peptide bonds
CC may be modified. For example, L-amino acids may be replaced by D-amino
CC acids, amine groups may be acetylated, and so on. The native antigenic
CC sequence and its antigenic derivatives may be used for detection of
CC hepatitis C virus and for raising antibodies against the virus.

XX SQ Sequence 44 AA;

Query Match 92.98; Score 143; DB 21; Length 44;

Best Local Similarity 92.98; Pred. No. 1.2e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKPPGGQIVG 28

Db 4 PKPQKTKRNTNRAPQDVKPPGGQIVG 31

Search completed: August 7, 2003, 11:14:07
Job time : 38.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-28
Perfect score: 154
Sequence: 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

al number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	93.5	3010	1 A45573	genome polyprotein
2	143	92.9	108	2 S41353	genome polyprotein
3	143	92.9	108	2 S41355	genome polyprotein
4	143	92.9	108	2 S41357	genome polyprotein
5	143	92.9	108	2 S41348	genome polyprotein
6	143	92.9	112	2 S41371	genome polyprotein
7	143	92.9	112	2 S41341	genome polyprotein
8	143	92.9	114	2 S41370	genome polyprotein
9	143	92.9	114	2 S41369	genome polyprotein
10	143	92.9	114	2 S41368	genome polyprotein
11	143	92.9	115	2 S41342	genome polyprotein
12	143	92.9	115	2 S41344	genome polyprotein
13	143	92.9	115	2 S41350	genome polyprotein
14	143	92.9	115	2 S41354	genome polyprotein
15	143	92.9	115	2 S41345	genome polyprotein
16	143	92.9	115	2 S41347	genome polyprotein
17	143	92.9	115	2 S41343	genome polyprotein
18	143	92.9	118	2 S41346	genome polyprotein
19	143	92.9	369	2 S21471	genome polyprotein
20	143	92.9	441	2 S12707	genome polyprotein
21	143	92.9	513	2 PC1284	genome polyprotein
22	143	92.9	520	2 JQ1925	polyprotein - hepa
23	143	92.9	523	2 JQ1926	polyprotein - hepa
24	143	92.9	550	2 JH0711	genome polyprotein
25	143	92.9	782	2 S19876	genome polyprotein
26	143	92.9	782	2 S18031	genome polyprotein
27	143	92.9	782	2 S18032	genome polyprotein
28	143	92.9	787	2 PN0677	hypothetical prote
29	143	92.9	874	2 JQ0883	genome polyprotein

30	143	92.9	874	2 JQ0881	genome polyprotein
31	143	92.9	876	2 PC2219	polypeptide - hepa
32	143	92.9	3010	1 GNVVTC	genome polyprotein
33	143	92.9	3010	1 GNVVTC	genome polyprotein
34	143	92.9	3010	1 S18030	genome polyprotein
35	143	92.9	3011	1 GNVVCH	genome polyprotein
36	143	92.9	3011	1 S40770	genome polyprotein
37	143	92.9	3014	1 JQ5620	genome polyprotein
38	143	92.9	3033	1 JQ1303	genome polyprotein
39	143	92.9	3033	1 GNVVJ8	genome polyprotein
40	141	91.6	88	2 S21336	genome polyprotein
41	139	90.3	108	2 S41356	genome polyprotein
42	139	90.3	114	2 S41359	genome polyprotein
43	139	90.3	114	2 S41358	genome polyprotein
44	139	90.3	115	2 S41351	genome polyprotein
45	139	90.3	115	2 S41349	genome polyprotein

ALIGNMENTS

RESULT 1

A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitisvirin (EC 3.4.21.98) (non)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hiji
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese cari
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; MID:9221612; PIDN:BAA01943.1; PID:922161
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 93.5%; Score 144; DB 1; Length 3010;
Best Local Similarity 92.9%; Pred. No. 1.1e-11;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28

Db 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 2

S41353
genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41353
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41353

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229456

A:Experimental source: genotype 2, N2

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 108;

Best Local Similarity 92.9%; Pred. No. 6.2e-13;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28

||||| :|||||

Db 5 PKPQKTKRNTNRRPQDVKFPGGGQIVG 32

JLT 3

--1355

genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 2, N4

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41355

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41355

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229458

A:Experimental source: genotype 2, N4

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 108;

Best Local Similarity 92.9%; Pred. No. 6.2e-13;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28

||||| :|||||

-- 5 PKPQKTKRNTNRRPQDVKFPGGGQIVG 32

RESULT 4

S41357

genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 2, N6

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41357

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41357

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229460

A:Experimental source: genotype 2, N6

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 108;

Best Local Similarity 92.9%; Pred. No. 6.2e-13;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28

||||| :|||||

Db 5 PKPQKTKRNTNRRPQDVKFPGGGQIVG 32

RESULT 5

S41348

genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 1, N6

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41348

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41348

A:Molecule type: genomic RNA

A:Residues: 1-108 <VAN>

A:Cross-references: EMBL:229451

A:Experimental source: genotype 1, N6

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-108/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 108;

Best Local Similarity 92.9%; Pred. No. 6.2e-13;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28

||||| :|||||

Db 5 PKPQKTKRNTNRRPQDVKFPGGGQIVG 32

RESULT 6

S41371

genome polyprotein - hepatitis C virus (genotype 5, N5) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 5, N5

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41371

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41371

A:Molecule type: genomic RNA

A:Residues: 1-112 <VAN>

A:Cross-references: EMBL:229474

A:Experimental source: genotype 5, N5

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-112/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 112;

Best Local Similarity 92.9%; Pred. No. 6.4e-13;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28

||||| :|||||

Db 5 PKPQKTKRNTNRRPQDVKFPGGGQIVG 32

RESULT 7

S41341

genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A: Variety: genotype 1, N1
C: Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C: Accession: S41341
R: van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
Submitted to the EMBL Data Library, January 1994
A: Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A: Reference number: S41341
A: Accession: S41341
A: Molecule type: genomic RNA
A: Residues: 1-112 <VAN>
A: Cross-references: EMBL:Z29444; NID:G443850; PIDN:CAA82582.1; PID:G443851
A: Experimental source: genotypel, N1
C: Superfamily: hepatitis C virus genome polyprotein
C: Keywords: capsid protein; core protein; polyprotein
F: 1-112/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 112;
Best Local Similarity 92.9%; Pred. No. 6.4e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

RESULT 8
S41370
genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)
N: Contains: core protein
C: Species: hepatitis C virus
A: Variety: genotype 5, N4
C: Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C: Accession: S41370
R: van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
Submitted to the EMBL Data Library, January 1994
A: Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A: Reference number: S41341
A: Accession: S41370
A: Molecule type: genomic RNA
A: Residues: 1-114 <VAN>
A: Cross-references: EMBL:Z29473; NID:G443908; PIDN:CAA82611.1; PID:G443909
A: Experimental source: genotype 5, N4
C: Superfamily: hepatitis C virus genome polyprotein
C: Keywords: capsid protein; core protein; polyprotein
F: 1-114/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 114;
Best Local Similarity 92.9%; Pred. No. 6.5e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

RESULT 9
S41369
genome polyprotein - hepatitis C virus (genotype 5, N3) (fragment)
N: Contains: core protein
C: Species: hepatitis C virus
A: Variety: genotype 5, N3
C: Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C: Accession: S41369
R: van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
Submitted to the EMBL Data Library, January 1994
A: Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A: Reference number: S41341
A: Accession: S41369
A: Molecule type: genomic RNA
A: Residues: 1-114 <VAN>
A: Cross-references: EMBL:Z29472; NID:G443906; PIDN:CAA82610.1; PID:G443907
A: Experimental source: genotype 5, N3
C: Superfamily: hepatitis C virus genome polyprotein

C: Keywords: capsid protein; core protein; polyprotein
F: 1-114/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 114;
Best Local Similarity 92.9%; Pred. No. 6.5e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

Db 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

RESULT 10
S41368
genome polyprotein - hepatitis C virus (genotype 5, N2) (fragment)
N: Contains: core protein
C: Species: hepatitis C virus
A: Variety: genotype 5, N2
C: Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C: Accession: S41368
R: van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
Submitted to the EMBL Data Library, January 1994
A: Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A: Reference number: S41341
A: Accession: S41368
A: Molecule type: genomic RNA
A: Residues: 1-114 <VAN>
A: Cross-references: EMBL:Z29471; NID:G443904; PIDN:CAA82609.1; PID:G443905
A: Experimental source: genotype 5, N2
C: Superfamily: hepatitis C virus genome polyprotein
C: Keywords: capsid protein; core protein; polyprotein
F: 1-114/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 114;
Best Local Similarity 92.9%; Pred. No. 6.5e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

Db 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

RESULT 11
S41342
genome polyprotein - hepatitis C virus (genotype 1, N10) (fragment)
N: Contains: core protein
C: Species: hepatitis C virus
A: Variety: genotype 1, N10
C: Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C: Accession: S41342
R: van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
Submitted to the EMBL Data Library, January 1994
A: Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A: Reference number: S41341
A: Accession: S41342
A: Molecule type: genomic RNA
A: Residues: 1-115 <VAN>
A: Cross-references: EMBL:Z29445; NID:G443852; PIDN:CAA82583.1; PID:G443853
A: Experimental source: genotype 1, N10
C: Superfamily: hepatitis C virus genome polyprotein
C: Keywords: capsid protein; core protein; polyprotein
F: 1-115/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 115;
Best Local Similarity 92.9%; Pred. No. 6.6e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFPFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRRRPQDVKFPFGGGQIVG 32

RESULT 12

S41344
genome polyprotein - hepatitis C virus (genotype 1, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41344
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41344
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229447; NID:9443856; PIDN:CAA82585.1; PID:9443857
A:Experimental source: genotype 1, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 115;
Best Local Similarity 92.9%; Pred. No. 6.6e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGOIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGGOIVG 32

RESULT 13
S41350
genome polyprotein - hepatitis C virus (genotype 1, N8) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N8
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41350
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41350
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229453; NID:9443868; PIDN:CAA82591.1; PID:9443869
A:Experimental source: genotype 1, N8
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 115;
Best Local Similarity 92.9%; Pred. No. 6.6e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGOIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGGOIVG 32

RESULT 14
S41354
genome polyprotein - hepatitis C virus (genotype 2, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41354
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41354
A:Molecule type: genomic RNA

A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229457
A:Experimental source: genotype 2, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 115;
Best Local Similarity 92.9%; Pred. No. 6.6e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGOIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGGOIVG 32

RESULT 15
S41345
genome polyprotein - hepatitis C virus (genotype 1, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41345
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41345
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:229448; NID:9443858; PIDN:CAA82586.1; PID:9443859
A:Experimental source: genotype 1, N3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 92.9%; Score 143; DB 2; Length 115;
Best Local Similarity 92.9%; Pred. No. 6.6e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGOIVG 28
DB 5 PKPQKTKRNTNRRPQDVKFPGGGOIVG 32

Search completed: August 7, 2003, 11:21:48
Job time : 9.54545 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds
(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146A-28
Perfect score: 154
Sequence: 1 PXPQRKTRNAHRPDVYKFGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
tal number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	93.5	3010	1 POLG_HCVJT	Q00269 h genome po
2	143	92.9	513	1 POLG_HCVJ2	P27959 hepatitis c
3	143	92.9	520	1 POLG_HCVH4	Q01404 hepatitis c
4	143	92.9	520	1 POLG_HCVHK	Q01403 hepatitis c
5	143	92.9	737	1 POLG_HCVJ5	P27960 hepatitis c
6	143	92.9	737	1 POLG_HCVJ7	P27961 hepatitis c
7	143	92.9	3010	1 POLG_HCVBK	P26663 h genome po
8	143	92.9	3010	1 POLG_HCVJA	P26662 h genome po
9	143	92.9	3011	1 POLG_HCVH	P27958 h genome po
10	143	92.9	3033	1 POLG_HCVJ6	P26660 h genome po
11	143	92.9	3033	1 POLG_HCVJ8	P26661 h genome po
12	136	88.3	3010	1 POLG_HCVTW	P29846 h genome po
13	135	87.7	3011	1 POLG_HCV1	P26564 h genome po
14	59	38.3	512	1 IE63_HSV2H	P28276 herpes simp
15	55	35.7	794	1 FURL_HUMAN	P09958 homo sapien
16	53	34.4	282	1 RK4_TOBAC	O80361 nicotiana t
17	50	32.5	194	1 RS7_FUGRU	P50894 fugu rubrip
18	50	32.5	194	1 RS7_HUMAN	P23821 homo sapien
19	50	32.5	293	1 RK4_SPIOL	O49937 spinacia ol
20	50	32.5	454	1 NCAP_CVM3	P18447 murine coro
21	50	32.5	454	1 NCAP_CVMA5	P03416 murine coro
22	50	32.5	454	1 NCAP_CVMS	P18448 murine coro
23	50	32.5	793	1 FURL_MOUSE	P23188 mus musculu
24	49	31.8	797	1 FURL_BOVIN	Q28193 bos taurus
25	48.5	31.5	303	1 YBL_XENLA	P21573 xenopus lae
26	48	31.2	317	1 CAH6_MOUSE	P18761 mus musculu
27	48	31.2	1531	1 YQ38_CAEEL	Q09459 caenorhabdi
28	48	31.2	1616	1 SLAP_BACCI	P35824 bacillus ci
29	47	30.5	190	1 RS7_MANSE	P48155 manduca sex
30	47	30.5	194	1 RS7_XENLA	P02362 xenopus lae
31	47	30.5	218	1 RIBB_VIBPA	Q87915 vibrio para
32	47	30.5	1311	1 FMR2_HUMAN	P51816 homo sapien
33	46	29.9	530	1 F262_BOVIN	P26285 bos taurus

RESULT 1
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID:31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:92295714; PubMed:1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RL infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D11168; BAA01943.1;
CC DR PIR: A45573; A45573.
CC DR PDB: 1A1Q; 25-MAR-98.
CC DR PDB: 1JXP; 14-JAN-98.
CC DR MEROPS: S29.001;
CC DR MEROPS: U39.001;
CC InterPro: IPR001410; DEAD.

015530 homo sapien
Q92240 mus musculu
O55173 rattus norv
P23377 rattus norv
O14738 h serine/th
P78317 homo sapien
Q03717 mus musculu
O95049 homo sapien
Q92444 mus musculu
O28653 o serine/th
P26786 saccharomyc
P48184 saccharomyc

[illegible]

FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FE27349B CRC64;

Query Match 92.9%; Score 143; DB 1; Length 520;
 Best Local Similarity 92.9%; Pred. No. 2.le-13;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDKVKGGOIGV 28
 ||||| :|||||
 Db 5 PKPQKTKRNTNRPPQDKVKGGOIGV 32

RESULT 5
 POLG_HCVJ5 STANDARD; PRT: 737 AA.
 AC P27960;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J5) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes."
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10075; BAA00969.1; .
 CC DR InterPro; IPR002522; HCV_capsid.
 CC DR InterPro; IPR002521; HCV_core.
 CC DR InterPro; IPR002519; HCV_env.
 CC DR InterPro; IPR002531; HCV_NS1.
 CC DR Pfam; PF01543; HCV_capsid; 1.
 CC DR Pfam; PF01542; HCV_core; 1.
 CC DR Pfam; PF01539; HCV_env; 1.
 CC DR Pfam; PF01560; HCV_NS1; 1.
 CC DR ProDom; PD186062; HCV_NS1; 1.
 CC DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
 CC Transmembrane; Nonstructural protein.
 KW INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT CHAIN NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT TRANSHEM 347 369
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 92.9%; Score 143; DB 1; Length 737;
 Best Local Similarity 92.9%; Pred. No. 3.le-13;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDKVKGGOIGV 28
 ||||| :|||||
 Db 5 PKPQKTKRNTNRPPQDKVKGGOIGV 32

RESULT 6
 POLG_HCVJ7 STANDARD; PRT: 737 AA.
 ID POLG_HCVJ7
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (isolate HC-J7) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes."
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10077; BAA00971.1; .
 CC DR InterPro; IPR002522; HCV_capsid.
 CC DR InterPro; IPR002521; HCV_core.
 CC DR InterPro; IPR002519; HCV_env.
 CC DR InterPro; IPR002531; HCV_NS1.
 CC DR Pfam; PF01543; HCV_capsid; 1.
 CC DR ProDom; PD186062; HCV_NS1; 1.
 CC DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;
 CC Transmembrane; Nonstructural protein.
 KW INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT CHAIN NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

DR DR Pfam: PF01542; HCV_core; 1.
DR DR Pfam: PF01539; HCV_env; 1.
DR DR Pfam: PF01560; HCV_NS1; 1.
DR DR ProDom: PD186062; HCV_NS1; 1.
KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein;
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 363 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 92.9%; Score 143; DB 1; Length 737;
Best Local Similarity 92.9%; Pred. No. 3.1e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTRNAHRPDQVKFPGGGQIVG 28
DB 5 PKPQKTRNTNRPDQVKFPGGGQIVG 32
|||||
|||||

POLG_HCVBK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
NT Genome polyprotein [Contains: Capsid protein c (Core protein) (P22);
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
from human carriers."
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits

```

```

RT phosphorylation mediated by cAMP-dependent protein kinase."
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomska Z.;
RA "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site."
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RA "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form."
RL Protein Sci. 7:837-847(1998).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M58335; AAA72945.1; .
DR PIR: A38465; GNVVTC.
DR PDB: 1A1Q; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR PDB: 1NS3; 08-APR-98.
DR PDB: 1C2P; 15-NOV-00.
DR PDB: 1CSJ; 08-NOV-99.
DR PDB: 1GX5; 09-APR-02.
DR PDB: 1GX6; 10-APR-02.
DR PDB: 1QOV; 26-JUN-00.
DR PDB: 8OHM; 20-APR-99.
DR MEROPS: S29.001; .
DR MEROPS: U39.001; .
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.

```


HSP: P26663; LUXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_Psvir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794FA6DB185 CRC64;
 Query Match 92.9%; Score 143; DB 1; Length 3010;
 Best Local Similarity 92.9%; Pred. No. 1.4e-12;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKYRNARRPQDVKFPGGQIVG 28
 ||||| :|||||
 Db 5 PKPQKTKRNTNRPPQDVKFPGGQIVG 32

RESULT 9

POLG_HCVH STANDARD; PRT; 3011 AA.
 ID AC 277958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:99-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

P26660;
 01-AUG-1992 (Rel. 23, Created)
 01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 Hepatitis C virus (isolate HC-J6) (HCV).
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 NCBI_TaxID=11113;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=9204440; PubMed=1658196;
 Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.,
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 J. Gen. Virol. 72:2697-2704(1991).
 -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 (RNA)(N).
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; D00944; BAA00792.1; -
 PIR; JQ1303; JQ1303.
 HSSP; P27958; 1HEI.
 MEROPS; S29.001; -
 MEROPS; U39.001; -
 InterPro; IPR001410; DEAD.
 InterPro; IPR002522; HCV_capsid.
 InterPro; IPR002521; HCV_core.
 InterPro; IPR002519; HCV_env.
 InterPro; IPR002531; HCV_NS1.
 InterPro; IPR002518; HCV_NS2.
 InterPro; IPR004109; HCV_NS3.
 InterPro; IPR000745; HCV_NS4A.
 InterPro; IPR001490; HCV_NS4B.
 InterPro; IPR002868; HCV_NS5A.
 InterPro; IPR002166; HCV_RdRP.
 InterPro; IPR001650; Helicase_C.
 InterPro; IPR007095; HCV_pol_DS_PS.
 InterPro; IPR007094; RNA_pol_PSVir.
 Pfam; PF01543; HCV_capsid; 1.
 Pfam; PF01542; HCV_core; 1.
 Pfam; PF01539; HCV_env; 1.
 Pfam; PF01560; HCV_NS1; 1.
 Pfam; PF01538; HCV_NS2; 1.
 Pfam; PF02907; HCV_NS3; 1.
 Pfam; PF01006; HCV_NS4a; 1.
 Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1234 1241 ATP (POTENTIAL).
 FT SITE 1320 1323 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 92.9%; Score 143; DB 1; Length 3033;
 Best Local Similarity 92.9%; Pred. No. 1.4e-12;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNAHRPQDVKPPGGQIVG 28
 Db 5 PKPQKTKRNTNRPPQDVKPPGGQIVG 32
 RESULT 11
 POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 ID POLG_HCVJ8
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE-9223032; PubMed-1314459;
RA Okamoto H., Kurai S., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes";
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10988; BAA01761.1;
DR PIR; A40250; GNMVJ8.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017

FT CHAIN 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;
Query Match 92.9%; Score 143; DB 1; Length 3033;
Best Local Similarity 92.9%; Pred. No. 1.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPORKTKRNARRRRPQDVKEFGGGIIVG 28
DB 5 PKPORKTKRNRRRPQDVKEFGGGIIVG 32
RESULT 12
POLG_HCVTW STANDARD; PRT; 3010 AA.
ID AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE Hepatitis C virus (isolate Taiwan) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230206; PubMed-1314449;
RA Chen P.-J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RT mapping the 5' termini of viral genomic and antigenomic RNA";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:


```

DR PIR: A39166; GNWVC3.
DR PDB: 1AIV; 16-FEB-99.
DR PDB: 1HEI; 25-NOV-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5a.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
DR PIR: A39166; GNWVC3.
DR PDB: 1AIV; 16-FEB-99.
DR PDB: 1HEI; 25-NOV-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5a.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE3AF9 CRC64;
Query Match 87.7%; Score 135; DB 1; Length 3011;
Best Local Similarity 85.7%; Pred. No. 2.1e-11;
Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQKTKRNAHRRPQDVKFPFGGQIVG 28
| | | | | | | | | | | | | | | | | |
DB 5 PKPQKKNKNTNRRPQDVKFPFGGQIVG 32
| | | | | | | | | | | | | | | | | |
RESULT 14
IE63_HSV2H
ID IE63_HSV2H STANDARD; PRT; 512 AA.
AC P28276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcriptional regulator IE63 (VNM63) (ICP27).
GN UL54.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID-10315;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92113549; PubMed-1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RT J. Gen. Virol. 72:3057-3075(1991).
RL [2]
RN SEQUENCE FROM N.A.
RP Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE DOWN-REGULATION OF VIRAL IMMEDIATE-EARLY
CC GENES AND ACTS IN COMBINATION WITH ICP0 AND ICP4 AS AN ACTIVATOR
CC OF LATE GENES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10471; BAA01269.1; -.
CC EMBL; 286099; CAB06702.1; -.
DR PIR: JQ1498; WMBEXA.
KW Early protein; Transcription regulation; Activator; DNA-binding.
SQ SEQUENCE 512 AA; 54958 MW; 459651470A503BA7 CRC64;
Query Match 38.3%; Score 59; DB 1; Length 512;
Best Local Similarity 40.6%; Pred. No. 0.47;
Matches 13; Conservative 3; Mismatches 12; Indels 4; Gaps 1;
QY 1 PKPQKTKRNAH-----RRPQDVKFPFGGQIVG 28
| | | | | | | | | | | | | | | | | |
DB 160 PKRRRVSRNAHNGRRHPASARTDGPCTATHG 191
| | | | | | | | | | | | | | | | | |
RESULT 15
FURI_HUMAN
ID FURI_HUMAN STANDARD; PRT; 794 AA.
AC P09958; Q14336;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Furin precursor (EC 3.4.21.75) (Paired basic amino acid residue

```

DE cleaving enzyme) (PACE) (Dibasic processing enzyme).

GN FURIN OR PACE OR FUR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Blood;

RX MEDLINE=90175002; PubMed=2408021;

RA van den Ouweland A.M.W., van Duijnhoven H.L.P., Kelzer G.D.,

RA Dorsers L.C.J., van de Ven W.J.M.;

RT "Structural homology between the human fur gene product and the

RT subtilisin-like protease encoded by yeast KEX2.";

RL Nucleic Acids Res. 18:664-664(1990).

RN [2]

RN SEQUENCE FROM N.A.

RX MEDLINE=91321735; PubMed=1713771;

RA Barr P.J., Mason O.B., Landsberg K.E., Wong P.A., Kiefer M.C.,

RA Brake A.J.;

RT "CDNA and gene structure for a human subtilisin-like protease with

RT cleavage specificity for paired basic amino acid residues.";

RL DNA Cell Biol. 10:319-328(1991).

RN [3]

RN SEQUENCE OF 296-794 FROM N.A.

RX MEDLINE=87053858; PubMed=3023061;

RA Roebroek A.J.M., Schaiken J.A., Leunissen J.A.M., Onnekink C.,

RA Bloemers H.P.J., van de Ven W.J.M.;

RT "Evolutionary conserved close linkage of the c-fes/tps proto-oncogene

RT and genetic sequences encoding a receptor-like protein.";

RL EMBO J. 5:2197-2202(1986).

RN [4]

RN 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.

RX MEDLINE=94291619; PubMed=8020465;

RA Slezen R.J., Creemers J.W.M., van de Ven W.J.M.;

RT "Homology modelling of the catalytic domain of human furin. A model

RT for the eukaryotic subtilisin-like proprotein convertases.";

RL Eur. J. Biochem. 222:255-266(1994).

RN [5]

RN PROCESSING.

RX MEDLINE=92332543; PubMed=1629222;

RA Leduc R., Molloy S.S., Thorne B.A., Thomas G.;

RT "Activation of human furin precursor processing endoprotease occurs by

RT an intramolecular autoproteolytic cleavage.";

RL J. Biol. Chem. 267:14304-14308(1992).

CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTIDASE

CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF

CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.

CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their

CC propeptides by cleavage of Arg-xaa-Arg-|-aa bonds, where xaa

CC can be any amino acid and yaa is Arg or Lys. Releases albumin,

CC complement component C3 and von Willebrand factor from their

CC respective precursors.

CC -1- COFACTOR: CALCIUM-DEPENDENT.

CC -1- ENZYME REGULATION: COULD BE INHIBITED BY THE NOT SECONDLY CLEAVED

CC PROPEPTIDE.

CC -1- SUBCELLULAR LOCATION: SEEMS TO BE LOCALIZED INTRACELLULARLY TO THE

CC TRANS GOLGI NETWORK. PROPEPTIDE CLEAVAGE IS A PREREQUISITE FOR

CC EXIT OF FURIN MOLECULES OUT OF THE ENDOPLASMIC RETICULUM (ER).

CC SECOND CLEAVAGE IN THE PROPEPTIDE OCCUR IN THE TRANS GOLGI NETWORK

CC (TGN), WHICH IS FOLLOWED BY THE RELEASE OF THE PROPEPTIDE BOUND TO

CC FURIN AND THE ACTIVATION OF FURIN.

CC -1- TISSUE SPECIFICITY: SEEMS TO BE EXPRESSED UBIQUITOUSLY.

CC -1- DOMAIN: CONTAINS A CYTOPLASMIC DOMAIN RESPONSIBLE FOR ITS TGN

CC LOCALIZATION AND RECYCLING FROM THE CELL SURFACE.

CC -1- PTM: THE PROPEPTIDE IS AUTOCATALYTICALLY REMOVED THROUGH AN

CC INTRAMOLECULAR CLEAVAGE PROBABLY IN THE ENDOPLASMIC RETICULUM

CC (ER). IN THE TGN THE SECOND CLEAVAGE IN THE PROPEPTIDE COULD LEAD

CC TO THE ACTIVATION OF FURIN.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.

CC -1- SIMILARITY: Contains 1 homo B/P domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X17094; CAA34948.1; -;

DR EMBL; X04329; CAA27860.1; -;

DR EMBL; A06939; CAA00605.1; -;

DR PIR; A39552; KXHUF.

DR HSP; Q99405; 1MPT.

DR MEROPS; S08.071; -;

DR Genew; HGNC:8568; FURIN.

DR MIM; 136950; -;

DR GO; GO:0005794; C:Golgi apparatus; TAS.

DR GO; GO:0004276; F:furin activity; TAS.

DR GO; GO:0007267; P:cell-cell signaling; TAS.

DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.

DR InterPro; IPR006212; Furin_repeat.

DR InterPro; IPR002884; P_domain.

DR Pfam; PF01483; P_protease; PARTIAL.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P_domain; 1.

DR SMART; SM00261; FU; 2.

DR PROSITE; PS00136; SUBTILASE_ASP; 1.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Serine protease; transmembrane; Glycoprotein; Signal;

KW Zymogen; Calcium.

FT SIGNAL 1 24 POTENTIAL.

FT PROPEP 25 107

FT CHAIN 108 794 FURIN.

FT DOMAIN 556 705 CYS-RICH.

FT TRANSHEM 716 738 POTENTIAL.

FT ACT_SITE 153 153 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 194 194 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SULFID 211 360 POTENTIAL.

FT DISULFID 303 333

FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SITE 70 75 CLEAVAGE (SECOND AUTO-).

FT SITE 104 107 CLEAVAGE (FIRST AUTO-).

FT SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 759 762 CELL SURFACE SIGNAL.

FT SITE 773 779 TRANS GOLGI NETWORK SIGNAL.

SQ SEQUENCE 794 AA; 86678 MW; 10C44DD5892EF35D CRC64;

Query Match 35.7%; Score 55; DB 1; Length 794;

Best Local Similarity 50.0%; Pred. No. 2.9;

Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 4 QKTKRNAHRPPDVKFP 21

Db 102 KRTKRDVYQETDPKFP 119

Search completed: August 7, 2003, 11:20:01

Job time : 4.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-28

Perfect score: 154

Sequence: 1 PKQRTKRNHRPQDVKPPGGQIVG 28

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	93.5	105	12 P90364	P90364 hepatitis c
2	144	93.5	106	12 Q81831	Q81831 hepatitis c
3	144	93.5	125	12 Q9PXN3	Q9PXN3 hepatitis c
4	144	93.5	191	12 Q68575	Q68575 hepatitis c
5	144	93.5	3010	12 Q91A00	Q91A00 hepatitis c
6	144	93.5	3010	12 Q81989	Q81989 hepatitis c
7	144	93.5	3010	12 Q81541	Q81541 hepatitis c
8	143	92.9	45	12 Q68311	Q68311 hepatitis c
9	143	92.9	45	12 Q68306	Q68306 hepatitis c
10	143	92.9	46	12 Q68309	Q68309 hepatitis c
11	143	92.9	60	12 Q8JYR9	Q8JYR9 hepatitis c
12	143	92.9	61	12 Q8JYR8	Q8JYR8 hepatitis c
13	143	92.9	61	12 Q8JYR5	Q8JYR5 hepatitis c
14	143	92.9	61	12 Q8JYR7	Q8JYR7 hepatitis c
15	143	92.9	62	12 Q8JYR6	Q8JYR6 hepatitis c
16	143	92.9	74	12 Q68712	Q68712 hepatitis c

17 143 92.9 74 12 Q68682 hepatitis c
18 143 92.9 74 12 Q68708 hepatitis c
19 143 92.9 74 12 Q68679 hepatitis c
20 143 92.9 74 12 Q68683 hepatitis c
21 143 92.9 74 12 Q68706 hepatitis c
22 143 92.9 74 12 Q68684 hepatitis c
23 143 92.9 74 12 Q68707 hepatitis c
24 143 92.9 74 12 Q68685 hepatitis c
25 143 92.9 76 12 Q8JY50 hepatitis c
26 143 92.9 78 12 Q8JY51 hepatitis c
27 143 92.9 83 12 Q81264 hepatitis c
28 143 92.9 100 12 Q9Q761 hepatitis c
29 143 92.9 100 12 Q8Q770 hepatitis c
30 143 92.9 100 12 Q8Q254 hepatitis c
31 143 92.9 100 12 Q9Q757 hepatitis c
32 143 92.9 100 12 Q8Q252 hepatitis c
33 143 92.9 100 12 Q8Q258 hepatitis c
34 143 92.9 100 12 Q9Q756 hepatitis c
35 143 92.9 100 12 Q8Q259 hepatitis c
36 143 92.9 100 12 Q8Q264 hepatitis c
37 143 92.9 100 12 Q8Q255 hepatitis c
38 143 92.9 100 12 Q8Q260 hepatitis c
39 143 92.9 100 12 Q8Q257 hepatitis c
40 143 92.9 100 12 Q9Q758 hepatitis c
41 143 92.9 100 12 Q8Q263 hepatitis c
42 143 92.9 100 12 Q8Q261 hepatitis c
43 143 92.9 100 12 Q9Q752 hepatitis c
44 143 92.9 100 12 Q9Q754 hepatitis c
45 143 92.9 100 12 Q8Q259 hepatitis c

ALIGNMENTS

RESULT 1
P90364 PRELIMINARY; PRT; 105 AA.
AC P90364;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohno T., Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR method."
RL Thesis (1995), Nagoya City University Medical School.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305314; PubMed=8712927;
RA Ohno T., Mizokami M., Saleh M.G., Orito E., Ohba K.-I., Wu R.-R., Koide T., Tibbs C.J., Nouri-Aria K.T., Tokudome S., Williams R.;
RT "Usefulness and limitation of phylogenetic analysis for hepatitis C virus core region: application to isolates from Egyptian and Yemeni patients";
RL Arch. Virol. 141:1101-1113(1996).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D82034; BAAL11519.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
FT NON_TER 1 105
FT NON_TER 105 105
SQ SEQUENCE 105 AA; 11870 MW; C7BA40B284025A49 CRC64;

```
Query Match          93.5%; Score 144; DB 12; Length 105;
Best Local Similarity 92.9%; Pred. No. 6.8e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28
DB 4 PKPQKTKRNTYRRPQDVKFPGGGQIVG 31

RESULT 2
Q81831 PRELIMINARY; PRT; 106 AA.
ID Q81831
AC Q81831
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE (Isolate USA8) genomic RNA (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
[1]
SEQUENCE FROM N.A.
RA STRAIN=USA8;
RA Ohno T., Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR
RT method.";
RL Submitted.;
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D49465; BAA08439.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 106 AA; 12001 MW; 25D0D5414B3EA9DC CRC64;

Query Match          93.5%; Score 144; DB 12; Length 106;
Best Local Similarity 92.9%; Pred. No. 6.9e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPQDVKFPGGGQIVG 32

RESULT 3
Q9PXN3 PRELIMINARY; PRT; 125 AA.
AC Q9PXN3
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE E2/NS1 protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus type 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=40271;
[1]
SEQUENCE FROM N.A.
RA MEDLINE=94351179; PubMed=7520922;
RA Nakazawa T., Kato N., Ohkoshi S., Shibuya A., Shimotohno K.;
RL J. Hepatol. 20:623-629(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
SQ SEQUENCE 125 AA; 14220 MW; D945CB60661797E3 CRC64;
```

```
Query Match          93.5%; Score 144; DB 12; Length 125;
Best Local Similarity 92.9%; Pred. No. 8.2e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPQDVKFPGGGQIVG 32

RESULT 4
Q68575 PRELIMINARY; PRT; 191 AA.
ID Q68575
AC Q68575
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
[1]
SEQUENCE FROM N.A.
RA STRAIN=BB51;
RA Songvilai S., Dharakul T., Kunkitti R., Thepthai C.;
RT "Molecular cloning and expression of hepatitis C virus core protein
RT and production of monoclonal antibodies to the recombinant protein.";
RL Asian Pac. J. Allergy Immunol. 14:0-0(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; U55284; AAB00215.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 191 AA; 20840 MW; 4AAE63444D8329E2 CRC64;

Query Match          93.5%; Score 144; DB 12; Length 191;
Best Local Similarity 92.9%; Pred. No. 1.3e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPQDVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPQDVKFPGGGQIVG 32

RESULT 5
Q91AU0 PRELIMINARY; PRT; 3010 AA.
ID Q91AU0
AC Q91AU0
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
[1]
SEQUENCE FROM N.A.
RA MEDLINE=21440119; PubMed=11556407;
RA Lim S.P., Khu Y.L., Hong W.J., Tay A., Ting A.E., Lim S.G., Tan Y.H.;
RT "Identification and molecular characterization of the complete genome
RT of a Singapore isolate of hepatitis C virus: sequence comparison with
RL other strains and phylogenetic analysis.";
RL Virus Genes 23:89-95(2001).
RN [2]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN-HCV-S1;
RA Lim S.P.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF356827; AAL00900.1; -.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01339; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS050507; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326793 MW; 3D89304314F9F795 CRC64;

Query Match 93.5%; Score 144; DB 12; Length 3010;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRPQDVKFGGGQIVG 28
|||||||:|||||||
5 PKPQRTKRNHRPQDVKFGGGQIVG 32

RESULT 6
Q81989 PRELIMINARY; PRT; 3010 AA.
AC Q81989;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE E1 and E2/NS1 envelope glycoprotein (Genome polyprotein).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-N;
RA Zheng W.-Y.Z.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-N;
RA Zheng W.Z.;
RT "Genotype identification of hepatitis c virus (HCV) isolated from a

```

```

RT single Japanese carrier in Nagasaki prefecture and genome analysis of
RT E1 and E2/NS1 envelope glycoprotein regions."
RL Jpn. J. Trop. Med. Hyg. 22:169-177(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D63857; BAA09919.1; -.
DR HSP6; P26863; IJXP.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR002165; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01339; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS050507; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327503 MW; C7BDB38169D6E3CF CRC64;

Query Match 93.5%; Score 144; DB 12; Length 3010;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRPQDVKFGGGQIVG 28
|||||||:|||||||
5 PKPQRTKRNHRPQDVKFGGGQIVG 32

Db

RESULT 7
Q81541 PRELIMINARY; PRT; 3010 AA.
AC Q81541;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JT;
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T.,
RA Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."

```

```

Virus Res. 23:39-53(1992).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; D11355; BAA18894.1; -
HSP; P26663; LJXP
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
Pfam; PF01543; HCV_capsid; 1.
Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_env; 1.
Pfam; PF01560; HCV_NS1; 1.
Pfam; PF01538; HCV_NS2; 1.
Pfam; PF02907; HCV_NS3; 1.
Pfam; PF01006; HCV_NS4a; 1.
Pfam; PF01001; HCV_NS4b; 1.
Pfam; PF01506; HCV_NS5a; 1.
Pfam; PF00998; Viral_RdRp; 1.
ProDom; PD186062; HCV_NS1; 1.
SMART; SM00487; DEXDC; 1.
PROSITE; PS05057; RDRP_POSITIVE; 1.
PROSITE; PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191 PUT. P22
FT CHAIN 192 383 PUT. GP35.
FT CHAIN 384 729 PUT. GP70.
FT CHAIN 730 1006 PUT. NS2.
FT CHAIN 1007 1615 PUT. NS3.
FT CHAIN 1616 1862 PUT. NS4A.
FT CHAIN 1863 2013 PUT. NS4B.
FT CHAIN 2014 3010 PUT. NS5.
SQ SEQUENCE 3010 AA; 326564 MW; 05F0B2102CF9DD9D CRC64;

Query Match 93.5%; Score 144; DB 12; Length 3010;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Caps 0;

1 PKPQRTKRNARRPQDVKPPGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRYRRPQDVKPPGGQIVG 32

RESULT 8
Q68311 ID Q68311 PRELIMINARY; PRT; 45 AA.
AC Q68311;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-HCV-BB37;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
```

```

Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23749; AAA65056.1; -
InterPro; IPR002522; HCV_capsid.
Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5015 MW; CC527167096AAA81 CRC64;

Query Match 92.9%; Score 143; DB 12; Length 45;
Best Local Similarity 92.9%; Pred. No. 3.9e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 PKPQRTKRNARRPQDVKPPGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRYRRPQDVKPPGGQIVG 32

Db

RESULT 9
Q68306 ID Q68306 PRELIMINARY; PRT; 45 AA.
AC Q68306;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-HCV-BB5;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
```

```

Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23749; AAA65056.1; -
InterPro; IPR002522; HCV_capsid.
Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 45 AA; 5030 MW; CC4C21ED236AAA81 CRC64;

Query Match 92.9%; Score 143; DB 12; Length 45;
Best Local Similarity 92.9%; Pred. No. 3.9e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 PKPQRTKRNARRPQDVKPPGGQIVG 28
||||| :|||||
5 PKPQRTKRNTRYRRPQDVKPPGGQIVG 32

Db

RESULT 10
Q68309 ID Q68309 PRELIMINARY; PRT; 46 AA.
AC Q68309;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (genome polyprotein) (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB10;
RA Sengsivilal S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
   virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: U23747; AAA65054.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 46 AA; 5129 MW; ABDCA4C21ED236AAA CRC64;

Query Match 92.9%; Score 143; DB 12; Length 46;
Best Local Similarity 92.9%; Pred. No. 4e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

./ 1 PKPQRTKRNHRRPQDVKFGGGQIVG 28
Db 5 PKPQRTKRNRRPQDVKFGGGQIVG 32

RESULT 11
Q8JYR9
ID Q8JYR9 PRELIMINARY; PRT; 60 AA.
AC Q8JYR9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG306;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: AF506616; AAM33392.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 60 AA; 6686 MW; B5691CB8F26F2F1 CRC64;

Query Match 92.9%; Score 143; DB 12; Length 60;
Best Local Similarity 92.9%; Pred. No. 5.3e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFGGGQIVG 28
Db 5 PKPQRTKRNRRPQDVKFGGGQIVG 32

RESULT 12
Q8JYR8
ID Q8JYR8 PRELIMINARY; PRT; 61 AA.
AC Q8JYR8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.

```

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG286;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: AF506617; AAM33393.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match 92.9%; Score 143; DB 12; Length 61;
Best Local Similarity 92.9%; Pred. No. 5.4e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFGGGQIVG 28
Db 5 PKPQRTKRNRRPQDVKFGGGQIVG 32

RESULT 13
Q8JYR5
ID Q8JYR5 PRELIMINARY; PRT; 61 AA.
AC Q8JYR5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG290;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
EMBL: AF506620; AAM33396.1; -.
InterPro: IPR002522; HCV_capsid.
Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 61 AA; 6897 MW; EC656DC79E8F26F2 CRC64;

Query Match 92.9%; Score 143; DB 12; Length 61;
Best Local Similarity 92.9%; Pred. No. 5.4e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNHRRPQDVKFGGGQIVG 28
Db 5 PKPQRTKRNRRPQDVKFGGGQIVG 32

RESULT 14
Q8JYR7
ID Q8JYR7 PRELIMINARY; PRT; 61 AA.
AC Q8JYR7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

```

Search completed: August 7, 2003, 11:19:01
Job time : 26.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-28

Perfect score: 154
Sequence: 1 PXPQRTKRNHRRPDVKFPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

tal number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	100.0	28	3	US-08-921-887-28
2	148	96.1	28	3	US-08-921-887-33
3	144	93.5	28	3	US-08-921-887-25
4	144	93.5	450	4	US-08-635-886C-191
5	144	93.5	450	4	US-08-635-886C-192
6	143	92.9	34	3	US-08-380-160-6
7	143	92.9	43	4	US-09-020-846-36
8	143	92.9	44	3	US-08-380-160-2
9	143	92.9	44	4	US-09-389-756-1
10	143	92.9	45	3	US-08-380-160-1
11	143	92.9	61	1	US-07-346-054-9
12	143	92.9	61	1	US-08-083-947-23
13	143	92.9	61	1	US-08-530-550-3
14	143	92.9	61	1	US-08-262-037-26
15	143	92.9	61	5	PCT-US93-08638-9
16	143	92.9	61	5	PCT-US94-07088-23
17	143	92.9	61	5	PCT-US95-13660-3
18	143	92.9	74	3	US-08-836-075A-10
19	143	92.9	74	4	US-08-635-886C-198
20	143	92.9	100	4	US-08-635-886C-232
21	143	92.9	108	3	US-08-836-075A-14
22	143	92.9	115	1	US-08-324-977-8
23	143	92.9	115	2	US-08-384-616-8
24	143	92.9	115	2	US-08-904-686A-8
25	143	92.9	115	3	US-09-315-850-8
26	143	92.9	123	2	US-08-501-195-2
27	143	92.9	137	3	US-08-836-075A-46

28 143 92.9 138 3 US-08-836-075A-60 Sequence 60, Appl
29 143 92.9 140 2 US-08-501-195-1 Sequence 1, Appl
30 143 92.9 154 3 US-08-854-531-2 Sequence 2, Appl
31 143 92.9 154 5 PCT-US95-13552-2 Sequence 2, Appl
32 143 92.9 190 1 US-07-681-701-16 Sequence 16, Appl
33 143 92.9 190 3 US-08-078-271B-1 Sequence 1, Appl
34 143 92.9 191 2 US-08-290-665A-155 Sequence 155, App
35 143 92.9 191 2 US-08-290-665A-156 Sequence 156, App
36 143 92.9 191 2 US-08-290-665A-157 Sequence 157, App
37 143 92.9 191 2 US-08-290-665A-158 Sequence 158, App
38 143 92.9 191 2 US-08-290-665A-159 Sequence 159, App
39 143 92.9 191 2 US-08-290-665A-160 Sequence 160, App
40 143 92.9 191 2 US-08-290-665A-161 Sequence 161, App
41 143 92.9 191 2 US-08-290-665A-163 Sequence 163, App
42 143 92.9 191 2 US-08-290-665A-164 Sequence 164, App
43 143 92.9 191 2 US-08-290-665A-165 Sequence 165, App
44 143 92.9 191 2 US-08-290-665A-166 Sequence 166, App
45 143 92.9 191 2 US-08-290-665A-167 Sequence 167, App

ALIGNMENTS

RESULT 1
US-08-921-887-28
; Sequence 28, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; METHOD OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SNA
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-28

Query Match 100.0%; Score 154; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 7.6e-16;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28
|||||
Db 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28

RESULT 2

US-08-921-887-33
; Sequence 33, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-33

Query Match 96.1%; Score 148; DB 3; Length 28;
Best Local Similarity 96.4%; Pred. No. 5.6e-15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28
|||||
Db 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28

RESULT 3

US-08-921-887-25
; Sequence 25, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:

Query Match 93.5%; Score 144; DB 4; Length 450;
Best Local Similarity 92.9%; Pred. No. 4e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-25

Query Match 93.5%; Score 144; DB 3; Length 28;
Best Local Similarity 92.9%; Pred. No. 2.1e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28
|||||
Db 1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28

RESULT 4

US-08-635-886C-191
; Sequence 191, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-191

Query Match 93.5%; Score 144; DB 4; Length 450;
Best Local Similarity 92.9%; Pred. No. 4e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPORKTKRNAHRRPQDVKFGGGQIVG 28
||||| :|||||
Db 5 PKPORKTKRNTYRRPQDVKFGGGQIVG 32

RESULT 5

US-08-635-886C-192
; Sequence 192, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; PRIOR FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 192
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-192

Query Match 93.5%; Score 144; DB 4; Length 450;
Best Local Similarity 92.9%; Pred. No. 4e-13;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPORKTKRNAHRRPQDVKFGGGQIVG 28
||||| :|||||
Db 5 PKPORKTKRNTYRRPQDVKFGGGQIVG 32

RESULT 6

US-08-380-160-6
; Sequence 6, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; STREET: OLIFF & BERRIDGE
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380.160
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057.471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30.024

; REFERENCE/DOCKET NUMBER: WPB 28682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-6400
; TELEFAX: (703)836-2787
; TELEX:

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
US-08-380-160-6

Query Match 92.9%; Score 143; DB 3; Length 34;
Best Local Similarity 92.9%; Pred. No. 3.7e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPORKTKRNAHRRPQDVKFGGGQIVG 28
||||| :|||||
Db 4 PKPORKTKRNTYRRPQDVKFGGGQIVG 31

RESULT 7

US-09-020-846-36
; Sequence 36, Application US/09020846
; Patent No. 6322965
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, Kenjiro
; APPLICANT: KASHIWAKUMA, Tomiko
; APPLICANT: CHIBA, Yukie
; APPLICANT: YAGI, Shintaro
; APPLICANT: HASEGAWA, Akira
; TITLE OF INVENTION: CHIMERA ANTIGEN PEPTIDE
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020.846
; FILING DATE: 09-FEB-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-027015
; FILING DATE: 10-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-024045
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 053466/0225
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-020-846-36

Query Match 92.9%; Score 143; DB 4; Length 43;
Best Local Similarity 92.9%; Pred. No. 4.7e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDVKFPGGGQIVG 28
DB 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 8

US-08-380-160-2
Sequence 2, Application US/08380160
Patent No. 6235284

GENERAL INFORMATION:

APPLICANT: DALBON, Pascal
APPLICANT: JOLIVET, Michel
TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
TITLE OF INVENTION: FOR DETECTING THE LATTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEFF & BERRIDGE
STREET: P.O. Box 19928
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/380,160
FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,471
FILING DATE: 06-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 28682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-6400
TELEFAX: (703)836-2787
TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Human Hepatitis C Virus
STRAIN: H77
FEATURE:

NAME/KEY: Peptide

LOCATION: 1..44

OTHER INFORMATION: /note= "N-terminal sequence of the
protein of the nucleocapsid or CORE protein of
the human hepatitis C virus"

US-08-380-160-2

Query Match 92.9%; Score 143; DB 3; Length 44;
Best Local Similarity 92.9%; Pred. No. 4.8e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDVKFPGGGQIVG 28
DB 4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31

RESULT 9

US-09-389-756-1
Sequence 1, Application US/09389756
Patent No. 6576240

GENERAL INFORMATION:

APPLICANT: JOLIVET, MICHEL
APPLICANT: PENIN, FRANCOIS
APPLICANT: DALBON, PASCAL
APPLICANT: LADAVIERE, LAURENT
APPLICANT: LACOUX, XAVIER
TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
TITLE OF INVENTION: TREATING AN HCV INFECTION
FILE REFERENCE: 103959
CURRENT APPLICATION NUMBER: US/09/389,756
CURRENT FILING DATE: 1999-09-07
EARLIER APPLICATION NUMBER: PCT/FR98/00442
EARLIER FILING DATE: 1998-03-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 44
TYPE: PRT
ORGANISM: Hepatitis C virus
PUBLICATION INFORMATION:
AUTHORS: Ogata, N. et al.
TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
Patent No. 6576240
TITLE: of Hepatitis Virus
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 3392-3396
DATE: 1991
RELEVANT RESIDUES: 2 TO 45

US-09-389-756-1

Query Match 92.9%; Score 143; DB 4; Length 44;
Best Local Similarity 92.9%; Pred. No. 4.8e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDVKFPGGGQIVG 28
DB 4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31

RESULT 10

US-08-380-160-1
Sequence 1, Application US/08380160
Patent No. 6235284

GENERAL INFORMATION:

APPLICANT: DALBON, Pascal
APPLICANT: JOLIVET, Michel
TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
TITLE OF INVENTION: FOR DETECTING THE LATTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEFF & BERRIDGE
STREET: P.O. Box 19928
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/380.160
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057.471
FILING DATE: 06-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 28682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-6400
TELEFAX: (703)836-2787
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Human Hepatitis C Virus
STRAIN: H77
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..45
OTHER INFORMATION: /note= "N-terminal sequence of the protein of the nucleocapside or CORE protein of the human hepatitis C virus"
US-08-380-160-1

Query Match 92.9%; Score 143; DB 3; Length 45;
Best Local Similarity 92.9%; Pred. No. 4.9e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNRHRRPQDVKFGGGQIVG 28
Db 5 PKPQKTKRNTNRHRRPQDVKFGGGQIVG 32

ULT 11
J-07-946-054-9
Sequence 9, Application US/07946054
Patent No. 5582968
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
APPLICANT: Hosein, Barbara H
TITLE OF INVENTION: No. 5582968el Branched Hybrid and Cluster
TITLE OF INVENTION: Peptides Effective in Diagnosing and Detecting No. 5582968-A,
TITLE OF INVENTION: No. 5582968-B Hepatitis
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: United Biomedical Inc.
STREET: 25 Davids Dr.
CITY: Hauppauge
STATE: New York
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/946.054
FILING DATE: 15-SEP-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34,045
REFERENCE/DOCKET NUMBER: 2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-273-2828
TELEFAX: 516-273-1717
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-946-054-9

Query Match 92.9%; Score 143; DB 1; Length 61;
Best Local Similarity 92.9%; Pred. No. 6.8e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNRHRRPQDVKFGGGQIVG 28
Db 4 PKPQKTKRNTNRHRRPQDVKFGGGQIVG 31

RESULT 12
US-08-083-947-23
Sequence 23, Application US/08083947
Patent No. 5639594
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
APPLICANT: Hosein, Barbara
TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
TITLE OF INVENTION: in Diagnosing and Detecting No. 5639594-A, No. 5639594-B
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: M. Lisa Wilson
STREET: 25 Davids Drive
CITY: Hauppauge
STATE: NY
COUNTRY: USA
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083.947
FILING DATE: 19930628
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 946.054
FILING DATE: 15-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34045
REFERENCE/DOCKET NUMBER: 2000Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-083-947-23

Query Match 92.9%; Score 143; DB 1; Length 61;
Best Local Similarity 92.9%; Pred. No. 6.8e-14;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDVKFPGGQIVG 28
 ||||| :|||||
 Db 4 PKPQKTKRNTNRRPDVKFPGGQIVG 31

RESULT 13

US-08-530-550-3
 ; Sequence 3, Application US/08530550
 ; Patent No. 5736321

GENERAL INFORMATION:
 APPLICANT: Hosein, Barbara
 APPLICANT: Wang, Chang Yi
 TITLE OF INVENTION: Peptides effective for Diagnosis and
 TITLE OF INVENTION: Detection of Hepatitis c Infection
 NUMBER OF SEQUENCES: 51
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: M. Lisa Wilson
 STREET: 25 Davids Drive
 CITY: Hauppauge
 STATE: NY
 COUNTRY: USA
 ZIP: 11788

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/530,550
 FILING DATE:

CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, M. Lisa
 REGISTRATION NUMBER: 34,045
 REFERENCE/DOCKET NUMBER: 20002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516)273-2828
 TELEFAX: (516)273-1717
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide

US-08-530-550-3

Query Match 92.9%; Score 143; DB 1; Length 61;
 Best Local Similarity 92.9%; Pred. No. 6.8e-14;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDVKFPGGQIVG 28
 ||||| :|||||
 Db 4 PKPQKTKRNTNRRPDVKFPGGQIVG 31

RESULT 14

US-08-262-037-26
 ; Sequence 26, Application US/08262037
 ; Patent No. 5747239

GENERAL INFORMATION:
 APPLICANT: Chang Yi Wang and Barbara Hosein
 TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
 TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
 TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
 NUMBER OF SEQUENCES: 136
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVE.
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP:

ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/262,037
 FILING DATE:

CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/719,819
 FILING DATE: 24-June-1991
 APPLICATION NUMBER: 07/667,275
 FILING DATE: 11-Mar-1991
 APPLICATION NUMBER: 07/651,735
 FILING DATE: 07-Feb-1991
 APPLICATION NUMBER: 07/538,799
 FILING DATE: 26-July-1990
 APPLICATION NUMBER: 07/510,153
 FILING DATE: 16-April-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323
 REFERENCE/DOCKET NUMBER: 1151-4043 US3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-758-4800
 TELEFAX: (212) 751-6849
 TELEX: 421792

INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: Amino acid
 STRANDEDNESS:
 TOPOLOGY: Unknown
 US-08-262-037-26

Query Match 92.9%; Score 143; DB 1; Length 61;
 Best Local Similarity 92.9%; Pred. No. 6.8e-14;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNAHRRPDVKFPGGQIVG 28
 ||||| :|||||
 Db 4 PKPQKTKRNTNRRPDVKFPGGQIVG 31

RESULT 15

PCT-US93-08638-9
 ; Sequence 9, Application PC/TUS9308638
 ; GENERAL INFORMATION:

APPLICANT: United Biomedical Inc.
 TITLE OF INVENTION: Novel Branched Hybrid and Cluster Peptides
 TITLE OF INVENTION: Effective in Diagnosing and Detecting Non-A,
 TITLE OF INVENTION: Non-B Hepatitis
 NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:
 ADDRESSEE: UNITED BIOMEDICAL INC.
 STREET: 25 Davids Drive
 CITY: Hauppauge
 STATE: New York
 COUNTRY: USA
 ZIP: 11788

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/08638
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: M. Lisa Wilson

;
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 9055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-2828
; TELEFAX: 516-273-1717
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-08638-9

Query Match 92.9%; Score 143; DB 5; Length 61;
Best Local Similarity 92.9%; Pred. No. 6.8e-14;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPORKTKRNAHRRPDVKFPGGGQIVG 28
4 PKPORKTKRNTNRRPDVKFPGGGQIVG 31

Search completed: August 7, 2003, 11:23:51
Job time : 10.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146a-28

Perfect score: 154

Sequence: 1 PKPQRTKNAHRPDVKEPGGGQIVG 28

Scoring table: BLOSUM62

Gapop.10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143	92.9	44	15	US-10-367-677-1
2	143	92.9	74	10	US-09-851-138-10
3	143	92.9	91	9	US-09-758-308-1
4	143	92.9	97	10	US-09-756-875-8
5	143	92.9	103	10	US-09-921-397-77
6	143	92.9	108	10	US-09-851-138-14
7	143	92.9	113	10	US-09-921-397-78
8	143	92.9	137	10	US-09-851-138-46
9	143	92.9	138	10	US-09-851-138-60
10	143	92.9	166	11	US-09-899-046-152
11	143	92.9	166	11	US-09-878-281-152
12	143	92.9	169	11	US-09-899-046-42
13	143	92.9	169	11	US-09-899-046-44
14	143	92.9	169	11	US-09-878-281-42
15	143	92.9	169	11	US-09-878-281-44

16	143	92.9	182	10	US-09-929-955-2
17	143	92.9	182	14	US-10-104-966-2
18	143	92.9	191	11	US-09-194-949-3
19	143	92.9	318	10	US-09-851-138-76
20	143	92.9	319	10	US-09-851-138-12
21	143	92.9	319	10	US-09-851-138-18
22	143	92.9	319	11	US-09-899-046-50
23	143	92.9	319	11	US-09-899-046-52
24	143	92.9	319	11	US-09-899-046-54
25	143	92.9	319	11	US-09-899-046-144
26	143	92.9	319	11	US-09-878-281-50
27	143	92.9	319	11	US-09-878-281-52
28	143	92.9	319	11	US-09-878-281-54
29	143	92.9	319	11	US-09-878-281-144
30	143	92.9	809	10	US-09-973-025-50
31	143	92.9	809	11	US-09-899-303-50
32	143	92.9	809	11	US-09-995-808-50
33	143	92.9	809	11	US-09-995-860-50
34	143	92.9	2894	10	US-09-941-611-23
35	143	92.9	2894	15	US-10-044-995-23
36	143	92.9	2985	15	US-10-259-275-40
37	143	92.9	3011	9	US-09-742-659-4
38	143	92.9	3011	10	US-09-952-572-9
39	143	92.9	3011	10	US-09-929-955-1
40	143	92.9	3011	10	US-09-747-419-20
41	143	92.9	3011	11	US-09-891-894-3
42	143	92.9	3011	14	US-10-104-966-1
43	143	92.9	3012	10	US-10-259-275-20
44	143	92.9	3012	10	US-09-238-076-2
45	143	92.9	3012	11	US-09-995-937-2

ALIGNMENTS

RESULT 1

US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10367677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; TITLE: of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match 92.9% Score 143: DB 15: Length 44;


```
; Best Local Similarity 92.9%; Pred. No. 5.9e-13; Indels 0; Gaps 0;
Matches 26; Conservative 1; Mismatches 1;

Qy 1 PKPQKTKRNAHRRPQDVKFGGQIVG 28
    ||||| :|||||
Db 4 PKPQKTKRNTNRRPQDVKFGGQIVG 31
    ||||| :|||||

RESULT 2
US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-851-138-10
Query Match 92.9%; Score 143; DB 10; Length 74;
Best Local Similarity 92.9%; Pred. No. 1e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFGGQIVG 28
    ||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32
    ||||| :|||||

RESULT 3
US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US20020090607A1
; GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS
; FILE REFERENCE: 14114.0349U2
; CURRENT APPLICATION NUMBER: US/09/758,308
```

```
; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 92.9%; Score 143; DB 9; Length 91;
Best Local Similarity 92.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFGGQIVG 28
    ||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32
    ||||| :|||||

RESULT 4
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 92.9%; Score 143; DB 10; Length 97;
Best Local Similarity 92.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFGGQIVG 28
    ||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32
    ||||| :|||||

RESULT 5
```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-851-138-14

Query Match          92.9%  Score 143;  DB 10;  Length 108;
Best Local Similarity 92.9%  Pred. No. 1.5e-12;
Matches 26;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

QY  1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28
      ||||| : |||||
Db   5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 7
US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: Sid nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-921-397-78

Query Match          92.9%  Score 143;  DB 10;  Length 113;
Best Local Similarity 92.9%  Pred. No. 1.6e-12;
Matches 26;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

QY  1 PKPQKTKRNAHRRPDQVKFPGGGQIVG 28
      ||||| : |||||
Db   5 PKPQKTKRNTNRRPDQVKFPGGGQIVG 32

RESULT 8
US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851.138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9

```

FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46
Query Match 92.9%; Score 143; DB 10; Length 137;
Best Local Similarity 92.9%; Pred. No. 1.9e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1 PKPQRTKRNHRPQDVKFGGGQIVG 28
||||| :|||||
5 PKPQRTKRNHRPQDVKFGGGQIVG 32
RESULT 9
US-09-851-138-60
Sequence 60, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-851-138-60
Query Match 92.9%; Score 143; DB 10; Length 138;
Best Local Similarity 92.9%; Pred. No. 2e-12;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQRTKRNHRPQDVKFGGGQIVG 28
||||| :|||||
DB 5 PKPQRTKRNHRPQDVKFGGGQIVG 32
RESULT 10
US-09-899-046-152
Sequence 152, Application US/09899046
Publication No. US2003008274A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-899-046-152
Query Match 92.9%; Score 143; DB 11; Length 166;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQRTKRNHRPQDVKFGGGQIVG 28
||||| :|||||
DB 5 PKPQRTKRNHRPQDVKFGGGQIVG 32
RESULT 11
US-09-878-281-152
Sequence 152, Application US/09878281
Publication No. US20030032005A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/878,281
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-878-281-152
Query Match 92.9%; Score 143; DB 11; Length 166;

Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28
||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 12

US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:

; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/899,046
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:

; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-42

Query Match 92.9%; Score 143; DB 11; Length 169;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28
||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 13

US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:

; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/899,046
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:

; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-44

Query Match 92.9%; Score 143; DB 11; Length 169;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28
||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 14

US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:

; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/878,281
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:

; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-42

Query Match 92.9%; Score 143; DB 11; Length 169;
Best Local Similarity 92.9%; Pred. No. 2.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNAHRRPQDVKFPGGQIVG 28
||||| :|||||
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 15

US-09-878-281-44
; Sequence 44, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:

; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/878,281
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:

; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-44

Query Match 92.9%; Score 143; DB 11; Length 169;
 Best Local Similarity 92.9%; Pred. No. 2.4e-12;
 Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNAHRRPQDVAFPGGQIVG 28
 Db ||||| :|||||
 5 PKPQKTKRNTNRRPQDVAFPGGQIVG 32

Search completed: August 7, 2003, 12:01:13
 Job time : 14.3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146A-27
Perfect score: 156
Sequence: 1 PKPQPKRNTPRRQDVKFGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
al number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	139	89.1	115	2 S41351	genome polyprotein
2	139	89.1	115	2 S41349	genome polyprotein
3	139	89.1	640	2 JQ1584	genome polyprotein
4	138	88.5	782	2 S19875	genome polyprotein
5	137	87.8	411	2 PC2060	genome polyprotein
6	136	87.2	88	2 S21336	genome polyprotein
7	135	86.5	108	2 S41353	genome polyprotein
8	135	86.5	108	2 S41355	genome polyprotein
9	135	86.5	108	2 S41357	genome polyprotein
10	135	86.5	108	2 S41348	genome polyprotein
11	135	86.5	112	2 S41371	genome polyprotein
12	135	86.5	112	2 S41341	genome polyprotein
13	135	86.5	114	2 S41370	genome polyprotein
14	135	86.5	114	2 S41369	genome polyprotein
15	135	86.5	114	2 S41368	genome polyprotein
16	135	86.5	115	2 S41342	genome polyprotein
17	135	86.5	115	2 S41344	genome polyprotein
18	135	86.5	115	2 S41350	genome polyprotein
19	135	86.5	115	2 S41354	genome polyprotein
20	135	86.5	115	2 S41345	genome polyprotein
21	135	86.5	115	2 S41347	genome polyprotein
22	135	86.5	115	2 S41343	genome polyprotein
23	135	86.5	118	2 S41346	genome polyprotein
24	135	86.5	369	2 S21471	genome polyprotein
25	135	86.5	441	2 S12707	genome polyprotein
26	135	86.5	513	2 PC1284	genome polyprotein
27	135	86.5	520	2 JQ1925	polyprotein - hepa
28	135	86.5	523	2 JQ1926	polyprotein - hepa
29	135	86.5	550	2 JH0711	genome polyprotein

30	135	86.5	782	2 S19876	genome polyprotein
31	135	86.5	782	2 S18031	genome polyprotein
32	135	86.5	782	2 S18032	genome polyprotein
33	135	86.5	787	2 PN0877	hypothetical prote
34	135	86.5	874	2 JQ0883	genome polyprotein
35	135	86.5	874	2 JQ0881	genome polyprotein
36	135	86.5	876	2 PC2219	polypeptide - hepa
37	135	86.5	3010	1 GNVVTC	genome polyprotein
38	135	86.5	3010	1 GNVVCT	genome polyprotein
39	135	86.5	3010	1 S18030	genome polyprotein
40	135	86.5	3011	1 GNVVCH	genome polyprotein
41	135	86.5	3011	1 S40770	genome polyprotein
42	135	86.5	3014	1 JC5620	genome polyprotein
43	135	86.5	3033	1 JQ1303	genome polyprotein
44	135	86.5	3033	1 GNVVJ8	genome polyprotein
45	134	85.9	108	2 S41356	genome polyprotein

ALIGNMENTS

RESULT 1

S41351
genome polyprotein - hepatitis C virus (genotype 1, N9) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N9
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41351
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41351
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:Z29454
A:Experimental source: genotype 1, N9
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.1%; Score 139; DB 2; Length 115;
Best Local Similarity 92.9%; Pred. No. 4.2e-11;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQPKRNTPRRQDVKFGGQIVG 28
||||| |||| ||||| ||||| |||||
Db 5 PKPQTKRNTNRPPQDVKFGGQIVG 32

RESULT 2

S41349
genome polyprotein - hepatitis C virus (genotype 1, N7) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N7
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41349
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41349
A:Molecule type: genomic RNA
A:Residues: 1-115 <VAN>
A:Cross-references: EMBL:Z29452; NID:g443866; PIDN:CAA82590.1; PID:g443867
A:Experimental source: genotype 1, N7
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-115/Product: core protein #status predicted <MAT>

Query Match 89.1%; Score 139; DB 2; Length 115;

```

Best Local Similarity 92.9%; Pred. No. 4.2e-11;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPQDVKFGGGQIVG 28
    ||||| |||| ||||| ||||| |||||
Db 5 PKPQRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 3
genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural protein
C:Species: hepatitis C virus
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 17-Nov-2000
C:Accession: JQ1584
R.Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
J. Gen. Virol. 73, 1521-1525, 1992
A:Title: Cloning and sequencing of the structural region and expression of putative core
A:Reference number: JQ1584; MUID:92300349; PMID:1318944
A:Accession: JQ1584
A:Molecule type: genomic RNA
A:Residues: 1-640 <KUM>
A:Cross-references: GB:X84079; NID:9643119; PIDN:CAA58888.1; PID:g643120
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro
F:1-191/Product: core protein C #status predicted <CPC>
F:192-383/Product: envelope protein E1 #status predicted <BE1>
F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted <
F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (cova
Query Match 89.1%; Score 139; DB 2; Length 640;
Best Local Similarity 92.9%; Pred. No. 2.2e-10;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPQDVKFGGGQIVG 28
    ||||| |||| ||||| ||||| |||||
Db 5 PKPQRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 4
genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK3
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: S19875
R.Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C virus iso
A:Reference number: S18029
A:Accession: S19875
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
A:Experimental source: isolate JK3
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>
Query Match 88.5%; Score 138; DB 2; Length 782;
Best Local Similarity 92.9%; Pred. No. 3.6e-10;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPQDVKFGGGQIVG 28
    ||||| |||| ||||| ||||| |||||
Db 5 PKPQRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 5

```

```

PC2060
genome polyprotein N1 - hepatitis C virus
N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2060
R.Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France
A:Reference number: PC2060; MUID:94197744; PMID:8147893
A:Accession: PC2060
A:Molecule type: mRNA
A:Residues: 1-411 <LI>
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural pro
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 87.8%; Score 137; DB 2; Length 411;
Best Local Similarity 89.3%; Pred. No. 2.6e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPQDVKFGGGQIVG 28
    ||||| |||| ||||| ||||| |||||
Db 5 PKPQRTKRNTNRRPQDVKFGGGQIVG 32

RESULT 6
S21336
genome polyprotein S29 (core protein region) - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
C:Accession: S21336
R.Sato, A.
submitted to the EMBL Data Library, April 1992
A:Description: A sensitive serodiagnosis of hepatitis C virus infection with two c
A:Reference number: S21336
A:Accession: S21336
A:Molecule type: genomic RNA
A:Residues: 1-88 <SAT>
A:Cross-references: EMBL:X65548; NID:g59492; PIDN:CAA46517.1; PID:g59493
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein
Query Match 87.2%; Score 136; DB 2; Length 88;
Best Local Similarity 89.3%; Pred. No. 7.9e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPQDVKFGGGQIVG 28
    ||||| |||| ||||| ||||| |||||
Db 11 PKPQRTKRNTNRRPQDVKFGGGQIVG 38

RESULT 7
S41353
genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41353
R.van Doorn, L.J.; Kletter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41353
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:229456
A:Experimental source: genotype 2, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein

```

```
F:1-108/Product: core protein #status predicted <MAT>

Query Match      86.5%; Score 135; DB 2; Length 108;
Best Local Similarity 89.3%; Pred. No. 1.3e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKQORQPKRNTPRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRRRPDQVKFPGGGQIVG 32

RESULT 8
S41355
genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41355
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41355
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:Z29458
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match      86.5%; Score 135; DB 2; Length 108;
Best Local Similarity 89.3%; Pred. No. 1.3e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKQORQPKRNTPRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRRRPDQVKFPGGGQIVG 32

RESULT 9
S41357
genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41357
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41357
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:Z29460
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match      86.5%; Score 135; DB 2; Length 108;
Best Local Similarity 89.3%; Pred. No. 1.3e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKQORQPKRNTPRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRRRPDQVKFPGGGQIVG 32

RESULT 10
S41348
genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41348
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41348
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:Z29451
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match      86.5%; Score 135; DB 2; Length 108;
Best Local Similarity 89.3%; Pred. No. 1.3e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKQORQPKRNTPRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRRRPDQVKFPGGGQIVG 32

RESULT 11
S41371
genome polyprotein - hepatitis C virus (genotype 5, N5) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N5
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41371
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41371
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
A:Cross-references: EMBL:Z29474
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>

Query Match      86.5%; Score 135; DB 2; Length 112;
Best Local Similarity 89.3%; Pred. No. 1.3e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKQORQPKRNTPRPDQVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRRRPDQVKFPGGGQIVG 32

RESULT 12
S41341
genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41341
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
A:Accession: S41341
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
```


A;Cross-references: EMBL:229444; NID:g443850; PIDN:CAA82582.1; PID:g443851
A;Experimental source: genotypel, N1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-112/Product: core protein #status predicted <MAT>

Query Match 86.5%; Score 135; DB 2; Length 112;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRPRQDVKFGGGQIVG 32

RESULT 13

S41370
genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 5, N4
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
A;Accession: S41370

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41370
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>

A;Cross-references: EMBL:229473; NID:g443908; PIDN:CAA82611.1; PID:g443909
A;Experimental source: genotype 5, N4
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 86.5%; Score 135; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRPRQDVKFGGGQIVG 32

RESULT 14

S41369
genome polyprotein - hepatitis C virus (genotype 5, N3) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 5, N3
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
A;Accession: S41369

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41369
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>

A;Cross-references: EMBL:229472; NID:g443906; PIDN:CAA82610.1; PID:g443907
A;Experimental source: genotype 5, N3
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 86.5%; Score 135; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRPRQDVKFGGGQIVG 32

RESULT 15

S41368
genome polyprotein - hepatitis C virus (genotype 5, N2) (fragment)
N;Contains: core protein
C;Species: hepatitis C virus
A;Variety: genotype 5, N2
C;Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
A;Accession: S41368

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A;Reference number: S41341

A;Accession: S41368
A;Molecule type: genomic RNA
A;Residues: 1-114 <VAN>

A;Cross-references: EMBL:229471; NID:g443904; PIDN:CAA82609.1; PID:g443905
A;Experimental source: genotype 5, N2
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; polyprotein
F;1-114/Product: core protein #status predicted <MAT>

Query Match 86.5%; Score 135; DB 2; Length 114;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRQDVKFGGGQIVG 28
|||||:|||||:|||||:|||||:|||||
Db 5 PKPQRTKNTNRPRQDVKFGGGQIVG 32

Search completed: August 7, 2003, 11:21:48
Job time : 10.5455 secs

RESULT 1	
POLG_HCVJ2	
ID	POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC	P27959;
DT	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE	Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE	(GP68) (GP70) (NS1) (Fragment)
DE	Hepatitis C virus (isolate HC-J2) (HCV).
OS	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepadnavirus.
OX	NCBI_TaxID=11111;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=922320232; PubMed=1314459;
RA	Okamoto H., Kuzai K., Okada S.I., Yamamoto K., Lizuka H.,
RA	Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT	*Full-length sequence of a hepatitis C virus genome having poor
RT	homology to reported isolates: comparative study of four distinct
RT	genotypes.*;
RL	Virology 188:331-341(1992).
CC	-1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC	PROTEIN C AND MRNA.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC	or send an email to license@lsb-sib.ch).
CC	-----
DR	EMBL: D10074; BAA00968.1;
DR	InterPro: IPR002522; HCV_capsid.
DR	InterPro: IPR002521; HCV_Core.
DR	InterPro: IPR002519; HCV_env.
DR	InterPro: IPR002531; HCV_NS1.
DR	Pfam: PF01543; HCV_capsid; 1.
DR	Pfam: PF01542; HCV_core; 1.
DR	Pfam: PF01539; HCV_env; 1.
DR	Pfam: PF01560; HCV_NS1; 1.
DR	ProDom: PD186062; HCV_NS1; 1.
DR	Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW	Transmembrane; Nonstructural protein.
FT	INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT	CELLULAR AMINOPEPTIDASE.
FT	CHAIN 1 115
FT	CHAIN 116 191
FT	CHAIN 192 383
FT	CHAIN 383
FT	MAJOR ENVELOPE PROTEIN E (POTENTIAL)
FT	MAJOR ENVELOPE PROTEIN E (POTENTIAL)

```
FT CHAIN 384 >513 NONSTRUCTURAL PROTEIN NSI (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 513 513
SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 86.5%; Score 135; DB 1; Length 513;
Best Local Similarity 89.3%; Pred. No. 1.3e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPKRNTPRRPQDVKFGGGQIVG 28
    |||||: ||||| ||||| ||||| |||||
5 PKPQKTKRNTNRRRPQDVKFGGGQIVG 32

RESULT 2
POLG_HCVH4
ID POLG_HCVH4 STANDARD; PRT: 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
DE Hepatitis C virus [Isolate HCV-476] (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC 1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10688; BAA01530.1;
CC InterPro; IPR002522; HCV capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002531; HCV_env.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NSI; 1.
CC ProDom; PD186062; HCV_NSI; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
```

```
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 >520 NONSTRUCTURAL PROTEIN NSI/E2 (POTENTIAL).
FT TRANSMEM 347 369 BY SIMILARITY.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 86.5%; Score 135; DB 1; Length 520;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPKRNTPRRPQDVKFGGGQIVG 28
    |||||: ||||| ||||| ||||| |||||
5 PKPQKTKRNTNRRRPQDVKFGGGQIVG 32

RESULT 3
POLG_HCVHK
ID POLG_HCVHK STANDARD; PRT: 520 AA.
AC Q01403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
DE Hepatitis C virus [Isolate HCV-KF] (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC 1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10687; BAA01529.1;
CC PIR; JQ1925; JQ1925.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NSI.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NSI; 1.
CC ProDom; PD186062; HCV_NSI; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
```

```

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 >520 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT TRANSMEM 347 369 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CARBOHYD 196 196 BY SIMILARITY.
FT CARBOHYD 209 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 520 AA; 1D2BD0A6FF27349B CRC64;

Query Match 86.5%; Score 135; DB 1; Length 520;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPORPKRNTPRRQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTNRPPQDVKPPGGQIVG 32

RESULT 4
POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RA *Full-length sequence of a hepatitis C virus genome having poor
RA homology to reported isolates: comparative study of four distinct
RA genotypes.*;
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10075; BAA00969.1; -
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.

```

```

DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 737 AA; 3AF699D82AD501B1 CRC64;

Query Match 86.5%; Score 135; DB 1; Length 737;
Best Local Similarity 89.3%; Pred. No. 2e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPORPKRNTPRRQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTNRPPQDVKPPGGQIVG 32

RESULT 5
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RA *Full-length sequence of a hepatitis C virus genome having poor
RA homology to reported isolates: comparative study of four distinct
RA genotypes.*;
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10075; BAA00969.1; -
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC Pfam: PF01543; HCV_capsid; 1.

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10077; BAA00971.1;
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR Pfam: PF01543; HCV_capsid.1.
DR Pfam: PF01542; HCV_core.1.
DR Pfam: PF01539; HCV_env.1.
DR Pfam: PF01560; HCV_NS1.1.
DR ProDom: PD186062; HCV_NS1.1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
CHAIN 1 115
CHAIN 116 191
CHAIN 192 383
CHAIN 384 733
CHAIN 734 >737
CHAIN TRSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 86.5%; Score 135; DB 1; Length 737;
Best Local Similarity 89.3%; Pred. No. 2e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

1 PKPQRPKNTPRRQDVKEFGGQIVG 28
|||||: ||||| ||||| ||||| |||||
5 PKPQRTKRNRRPDQVKEFGGQIVG 32

RESULT 6
POLG_HCVBK STANDARD; PRT: 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11105;
OX

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-91140698; PubMed-1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.:
RA "Structure and organization of the hepatitis C virus genome isolated
RA from human carriers.";
RA J. Virol. 65:1103-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RA MEDLINE-9623524; PubMed-8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
RA "Non-structural protein 3 of hepatitis C virus inhibits
RA phosphorylation mediated by cAMP-dependent protein kinase.";
RA Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RA MEDLINE-97015088; PubMed-8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomska Z.;
RA "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RA trypsin-like fold and a structural zinc binding site.";
RA Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RA MEDLINE-98227846; PubMed-9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RA "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RA virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RA Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M58335; AAA72945.1;
DR PIR: A38465; GNVVTC.
DR PDB: 1A10; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR PDB: 1NS3; 08-APR-98.
DR PDB: 1C2P; 15-NOV-00.
DR PDB: 1G5J; 08-APR-02.
DR PDB: 1GX5; 09-APR-02.
DR PDB: 1CX6; 10-APR-02.
DR PDB: 1QVJ; 26-JUN-00.
DR PDB: 80HM; 20-APR-99.
DR MEROPS: S29.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.

```

DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.
 INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN TRANSMEM 347 369
 CHAIN ACT_SITE 1083 1083
 CHAIN ACT_SITE 1107 1107
 CHAIN ACT_SITE 1165 1165
 CHAIN NP_BIND 1230 1237
 CHAIN SITE 1316 1319
 CHAIN CARBOHYD 196 196
 CHAIN CARBOHYD 209 209
 CHAIN CARBOHYD 234 234
 CHAIN CARBOHYD 250 250
 CHAIN CARBOHYD 305 305
 CHAIN CARBOHYD 417 417
 CHAIN CARBOHYD 423 423
 CHAIN CARBOHYD 430 430
 CHAIN CARBOHYD 448 448
 CHAIN CARBOHYD 532 532
 CHAIN CARBOHYD 540 540
 CHAIN CARBOHYD 556 556
 CHAIN CARBOHYD 576 576
 CHAIN CARBOHYD 623 623
 CHAIN CARBOHYD 645 645
 CHAIN CARBOHYD 2041 2041
 CHAIN CARBOHYD 2077 2077
 CHAIN CARBOHYD 2240 2240
 CHAIN CARBOHYD 2529 2529
 CHAIN CARBOHYD 2788 2788
 CHAIN STRAND 1031 1035
 CHAIN HELIX 1039 1047
 CHAIN STRAND 1050 1050
 CHAIN STRAND 1059 1063
 CHAIN STRAND 1068 1074
 CHAIN TURN 1075 1076
 CHAIN STRAND 1077 1081
 CHAIN HELIX 1082 1085
 CHAIN TURN 1086 1087
 CHAIN STRAND 1090 1092
 CHAIN TURN 1093 1094
 CHAIN STRAND 1095 1097
 CHAIN STRAND 1101 1103
 CHAIN TURN 1104 1107

FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT STRAND 1165 1166
 FT TURN 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;
 Query Match 86.5%; Score 135; DB 1; Length 3010;
 Best Local Similarity 89.3%; Pred. No. 8.9e-10;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQRPKRNTPRPQDVKEPFGGQIVG 28
 DB 5 PKPQRTKRNTRRPQDVKEPFGGQIVG 32
 RESULT 7
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RA "Molecular cloning of the human hepatitis C virus genome from
 RA Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 [2]
 RN DISCUSSION OF SEQUENCE.
 RP MEDLINE=91192160; PubMed=1849488;
 RX Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
 RA Ohkoshi S., Shimotohno K.;
 RL "Molecular structure of the Japanese hepatitis C viral genome.";
 FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M67463; AAA45534.1; -
CC PIR: A36814; GNMVCH.
CC PDB: 1HEI; 25-NOV-98.
CC PDB: 1AIV; 16-FEB-99.
CC PDB: 1AIR; 17-JUN-98.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC TRANSFAC: T04155; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRp.
CC InterPro: IPR001650; Helicase_C.
CC InterPro: IPR007095; RNA_pol_DS_PS.
CC InterPro: IPR007094; RNA_pol_PSVir.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00271; helicase_C; 1.
CC Pfam: PF00998; Viral_RdRp; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC SMART: SM00487; DEXDC; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
CC 3D-structure.
CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CC CELLULAR AMINOPEPTIDASE.
CC CHAIN 1 191
CC CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
CC CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
CC CHAIN 747 809 PROTEIN P7.
CC CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
CC CHAIN 1027 1657 PROTEASE/HELICASE NS3.
CC CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
CC CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
CC CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
CC CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
CC CHAIN 347 369 POTENTIAL.
CC TRANSMEM 347 369
CC ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC NP_BIND 1230 1237 ATP (POTENTIAL).
CC SITE 1316 1319 DECH BOX.

FT CARBOHYD 196
FT CARBOHYD 209
FT CARBOHYD 224
FT CARBOHYD 305
FT CARBOHYD 417
FT CARBOHYD 423
FT CARBOHYD 430
FT CARBOHYD 438
FT CARBOHYD 476
FT CARBOHYD 532
FT CARBOHYD 540
FT CARBOHYD 556
FT CARBOHYD 576
FT CARBOHYD 623
FT CARBOHYD 645
FT STRAND 1224
FT TURN 1232
FT TURN 1233
FT TURN 1236
FT TURN 1238
FT HELIX 1239
FT TURN 1246
FT TURN 1247
FT STRAND 1248
FT HELIX 1251
FT HELIX 1258
FT TURN 1272
FT STRAND 1272
FT TURN 1280
FT TURN 1281
FT STRAND 1282
FT STRAND 1283
FT STRAND 1285
FT STRAND 1291
FT HELIX 1295
FT TURN 1301
FT TURN 1302
FT TURN 1303
FT TURN 1312
FT TURN 1316
FT TURN 1317
FT HELIX 1319
FT TURN 1323
FT TURN 1335
FT TURN 1336
FT STRAND 1340
FT STRAND 1343
FT TURN 1347
FT TURN 1352
FT TURN 1353
FT TURN 1360
FT STRAND 1361
FT STRAND 1362
FT STRAND 1366
FT STRAND 1368
FT STRAND 1373
FT TURN 1375
FT TURN 1376
FT STRAND 1377
FT STRAND 1380
FT HELIX 1382
FT STRAND 1385
FT STRAND 1389
FT HELIX 1393
FT TURN 1397
FT TURN 1401
FT STRAND 1410
FT TURN 1411
FT TURN 1417
FT TURN 1419
FT STRAND 1420
FT STRAND 1432
FT TURN 1436
FT TURN 1438
FT STRAND 1439
FT STRAND 1450
FT STRAND 1453
FT STRAND 1456
FT STRAND 1471
FT STRAND 1478
FT STRAND 1480
FT HELIX 1481
FT TURN 1488
FT TURN 1490
FT STRAND 1497
FT STRAND 1501
FT STRAND 1507
FT STRAND 1511
FT HELIX 1511
FT HELIX 1514
FT HELIX 1527
FT STRAND 1532
FT STRAND 1544
FT STRAND 1550
FT HELIX 1555
FT HELIX 1564
FT HELIX 1570
FT TURN 1578
FT TURN 1580
FT HELIX 1584
FT TURN 1597
FT TURN 1598
FT TURN 1606
FT HELIX 1611
FT TURN 1614
FT STRAND 1618
FT STRAND 1622
FT STRAND 1623
FT STRAND 1627
FT STRAND 1635
FT STRAND 1636
FT HELIX 1640
FT SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;


```

Query Match      86.5%; Score 135; DB 1; Length 3011;
Best Local Similarity 89.3%; Pred. No. 8.9e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPDQVKFPGGQIVG 28
    |||||: ||||| ||||| ||||| |||||
DB 5 PKPQRTKRNTRRPQDVKFPGGQIVG 32

RESULT 9
POLG_HCVJ6      STANDARD; PRT; 3033 AA.
ID POLG_HCVJ6
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP33) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (Isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11113;
RN [1]
RP SEQUENCE FROM N.A. PubMed=1658196;
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions".
RL J. Gen. Virol. 72:2697-2704(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1',
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00944; BAA00792.1; -
CC PIR; J01303; J01303.
CC HSSP; P27958.1HEI.
CC MEROPS; S29.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.

```

```

DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_NS5a.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01542; HCV_capsid; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match      86.5%; Score 135; DB 1; Length 3033;
Best Local Similarity 89.3%; Pred. No. 9e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPDQVKFPGGQIVG 28
    |||||: ||||| ||||| ||||| |||||
DB 5 PKPQRTKRNTRRPQDVKFPGGQIVG 32

RESULT 10
POLG_HCVJ8      STANDARD; PRT; 3033 AA.
ID POLG_HCVJ8
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

```


RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals."
 RL Virus Res. 23:39-53(1992).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D11168; BAA01843.1; -
 DR PIR: A45573; A45573.
 DR PDB: 1A1Q; 25-MAR-98.
 DR PDB: 1JXP; 14-JAN-98.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 85.9%; Score 134; DB 1; Length 3010;
 Best Local Similarity 89.3%; Pred. No. 1.2e-09;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 PKPQPKRNTPRPQDVKPGGQIVG 28
 Db 5 PKPQKTKRTYRPPQDVKPGGQIVG 32
 RESULT 12
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 ID POLG_HCV1
 AC P26654;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11104;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Welner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus."
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

DR PIR: A40244; GNNVTW.
 DR PDB: 1N64; 25-FEB-03.
 DR PDB: 1NS3; 08-APR-98.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 FT CHAIN 1115
 FT CHAIN 1191
 FT CHAIN 192
 FT CHAIN 384
 FT CHAIN 730
 FT CHAIN 1007
 FT CHAIN 1616
 FT CHAIN 1863
 FT CHAIN 2014
 FT TRANSMEM 347
 FT ACT_SITE 1083
 FT ACT_SITE 1107
 FT ACT_SITE 1165
 FT NP_BIND 1230
 FT SITE 1316
 FT CARBOHYD 196
 FT CARBOHYD 209
 FT CARBOHYD 233
 FT CARBOHYD 234
 FT CARBOHYD 250
 FT CARBOHYD 305
 FT CARBOHYD 417
 FT CARBOHYD 423
 FT CARBOHYD 430
 FT CARBOHYD 448
 FT CARBOHYD 532
 FT CARBOHYD 540
 FT CARBOHYD 556
 FT CARBOHYD 576
 FT CARBOHYD 623
 FT CARBOHYD 645
 FT CARBOHYD 2041
 FT CARBOHYD 2077
 FT CARBOHYD 2240
 FT CARBOHYD 2529
 FT CARBOHYD 2788
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 82.1%; Score 128; DB 1; Length 3010;
 Best Local Similarity 88.9%; Pred. No. 7.6e-09;
 Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 2 KFORQPKRNTPRRQDVKPPGGQIVG 28
 DB 6 KFORQPKRNTPRRQDVKPPGGQIVG 32
 RESULT 14
 FMR2_HUMAN
 ID FMR2_HUMAN STANDARD; PRT; 1311 AA.
 AC P51816; O43786; O60215; P78407; Q13521; Q14323; Q9UNAS5;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Fragile X mental retardation 2 protein (Protein FMR-2) (Ox19
 GN protein) (Fragile X E mental retardation syndrome protein).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Fetal brain, and Placenta;
 RX MEDLINE=96241584; PubMed=8673086;
 RA Geiz J., dedon A.K., Sutherland G.R., Mulley J.C.;
 RT "Identification of the gene FMR2, associated with FRAXE mental
 RL Nat. Genet. 13:105-108(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=96241584; PubMed=8673086;
 RA Gu Y., Shen Y., Gibbs R.A., Nelson D.L.;
 RT "Identification of FMR2, a novel gene associated with the FRAXE CCG
 RL repeat and Cpg island.";
 RN Nat. Genet. 13:109-113(1996).
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Fetal brain;
 RX MEDLINE=9642267; PubMed=8824884;
 RA Chakrabarti L., Knight S.J.L., Flannery A.V., Davies K.E.;
 RT "A candidate gene for mild mental handicap at the FRAXE fragile
 RL site.";
 RN Hum. Mol. Genet. 5:275-282(1996).
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=97446139; PubMed=9299237;
 RA Geiz J., Bielby S., Sutherland G.R., Mulley J.C.;
 RT "Gene structure and subcellular localization of FMR2, a member of a
 RL new family of putative transcription activators.";
 RN Genomics 44:201-213(1997).
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=98133924; PubMed=9467002;
 RA Chakrabarti L., Bristulif J., Foss G.S., Davies K.E.;
 RT "Expression of the murine homologue of FMR2 in mouse brain and during
 RL development.";
 RN Hum. Mol. Genet. 7:441-448(1998).
 RP SEQUENCE OF 520-1311 FROM N.A. (ISOFORM 3).
 RA Dugan-Rocha S.P., Gibbs R.A.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE OF 348-421 FROM N.A. (ISOFORM 1).
 RA Wang L., Thibodeau S.N.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- ALTERNATIVE PRODUCTS:


```

FT DISULFID 364 433 BY SIMILARITY.
FT DISULFID 368 435 BY SIMILARITY.
FT DISULFID 400 400 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 436 AA; 47873 MW; B0688E12EF8AE91D CRC64;

Query Match 39.7%; Score 62; DB 1; Length 436;
Best Local Similarity 41.7%; Pred. No. 0.56;
Matches 10; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPODVKPPGGG 24
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 41 PRQEEPRRPPQOPEAREPPGPG 64

```

Search completed: August 7, 2003, 11:20:01
 Job time : 6.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-27

Perfect score: 156

Sequence: 1 PKQRPQRNTPRPQDVKPPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

tal number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL23:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_virus:**
- 16: sp_bacteriaph:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	415	12	Q81548 hepatitis c
2	144	92.3	119	12	Q8BCW3 hepatitis c
3	143	91.7	109	12	Q81230 hepatitis c
4	143	91.7	109	12	Q81235 hepatitis c
5	143	91.7	109	12	Q81233 hepatitis c
6	143	91.7	109	12	Q81231 hepatitis c
7	143	91.7	109	12	Q81232 hepatitis c
8	143	91.7	109	12	Q81234 hepatitis c
9	143	91.7	109	12	Q81236 hepatitis c
10	143	91.7	150	12	Q68867 hepatitis c
11	143	91.7	150	12	Q68861 hepatitis c
12	143	91.7	150	12	Q68863 hepatitis c
13	143	91.7	150	12	Q68865 hepatitis c
14	143	91.7	191	12	Q8V7S2 hepatitis c
15	143	91.7	191	12	Q8V7T0 hepatitis c
16	143	91.7	191	12	Q8V7S4 hepatitis c

17	143	91.7	416	12	Q81265
18	139	89.1	74	12	Q68880
19	139	89.1	100	12	Q8QP62
20	139	89.1	114	12	Q8QM76
21	139	89.1	115	12	Q68883
22	139	89.1	116	12	Q68885
23	139	89.1	119	12	Q81486
24	139	89.1	119	12	Q8JP06
25	139	89.1	128	12	Q81542
26	139	89.1	191	12	Q8JWK6
27	139	89.1	191	12	Q8JWK0
28	139	89.1	191	12	Q8V7J4
29	139	89.1	191	12	Q70635
30	139	89.1	191	12	Q8JWM0
31	139	89.1	191	12	Q8JWK7
32	139	89.1	191	12	Q8JWK1
33	139	89.1	191	12	Q68126
34	139	89.1	414	12	Q68796
35	139	89.1	415	12	Q81266
36	139	89.1	640	12	Q68966
37	139	89.1	3010	12	P88803
38	139	89.1	3011	12	Q9DIT6
39	139	89.1	3022	12	Q68798
40	138	88.5	43	12	Q68305
41	138	88.5	109	12	Q81807
42	138	88.5	782	12	Q68951
43	137	87.8	802	12	Q68520
44	137	87.8	3033	12	Q991B5
45	136	87.2	88	12	Q68967

ALIGNMENTS

RESULT 1

Q81548
ID Q81548 PRELIMINARY; PRT; 415 AA.
AC Q81548;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core, env, and part of E2/NS1 (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-NEI25;
RX MEDLINE=94186155; PubMed=8138250;
RA Shrestha S.M., Tsuda F., Okamoto H., Tokita H., Horikita M., Tanaka T., Miyakawa Y., Mayumi M.;
RT "Hepatitis B virus subtypes and hepatitis C virus genotypes in patients with chronic liver disease in Nepal.";
RL Hepatology 19:805-809(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NEI25;
RX MEDLINE=94201770; PubMed=8151307;
RA Tokita H., Shrestha S.M., Okamoto H., Sakamoto M., Horikita M., Iizuka H., Shrestha S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group.";
RL J. Gen. Virol. 75:931-936(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D16614; BAA04036.1; -;
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NSL.

Q81265 hepatitis c
Q68880 hepatitis c
Q8QP62 hepatitis c
Q8QM76 hepatitis c
Q68883 hepatitis c
Q68885 hepatitis c
Q81486 hepatitis c
Q8JP06 hepatitis c
Q81542 hepatitis c
Q8JWK6 hepatitis c
Q8JWK0 hepatitis c
Q8V7J4 hepatitis c
Q70635 hepatitis c
Q8JWM0 hepatitis c
Q8JWK7 hepatitis c
Q8JWK1 hepatitis c
Q68126 hepatitis c
Q68796 hepatitis c
Q81266 hepatitis c
Q68966 hepatitis c
P88803 hepatitis c
Q9DIT6 hepatitis c
Q68798 hepatitis c
Q68305 hepatitis c
Q81807 hepatitis c
Q68951 hepatitis c
Q68520 hepatitis c
Q991B5 hepatitis c
Q68967 hepatitis c


```

DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 415 415
SQ SEQUENCE 415 AA; 44859 MW; 98C0EFDE22B891B9 CRC64;

Query Match
Best Local Similarity 100.0%; Score 156; DB 12; Length 415;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
DB 5 PKPQPKRNTPRPDQVKFPGGGQIVG 32

RESULT 2
Q8BCW3 PRELIMINARY; PRT; 119 AA.
AC Q8BCW3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=061653a;
RA Theamboonlers A., Bedi K., Chinchai T., Sriponthong M.,
RA Chantarasamee P., Poovorawan Y.;
RT "Molecular characterization of HCV core region of HCV infection in
RT Thai blood donor";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RE EMBL: AF525905; AAI16428.1; -
FT NON_TER 119 119
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 13427 MW; 0EBB321B6CB26A0A CRC64;

Query Match
Best Local Similarity 92.3%; Score 144; DB 12; Length 119;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
1 PKPQPKRNTPRPDQVKFPGGGQIVG 28

RESULT 3
Q81230 PRELIMINARY; PRT; 109 AA.
AC Q81230;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bani-1;
RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;
RT "The nucleotide sequence of the core region of HCV in Pakistan and
RT Bangladesh: The geographic characterization of HCV in South Asia.";
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D29652; BAA21027.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12235 MW; D5D1BBB10A7E3CAB CRC64;

Query Match
Best Local Similarity 91.7%; Score 143; DB 12; Length 109;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
5 PKPQPKRNTPRPDQVKFPGGGQIVG 32

RESULT 4
Q81235 PRELIMINARY; PRT; 109 AA.
AC Q81235;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ban2-3;
RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;
RT "The nucleotide sequence of the core region of HCV in Pakistan and
RT Bangladesh: The geographic characterization of HCV in South Asia.";
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D29652; BAA21027.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12235 MW; D5D1BBB10A7E3CAB CRC64;

Query Match
Best Local Similarity 92.9%; Score 143; DB 12; Length 109;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
5 PKPQPKRNTPRPDQVKFPGGGQIVG 32

RESULT 5
Q81233 PRELIMINARY; PRT; 109 AA.
AC Q81233;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ban2-1;
RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;

```

```

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D29647; BAA21022.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12293 MW; 524E8D425A7E3CH2 CRC64;

Query Match
Best Local Similarity 91.7%; Score 143; DB 12; Length 109;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
DB 5 PKPQPKRNTPRPDQVKFPGGGQIVG 32

```

```

RESULT 4
Q81235 PRELIMINARY; PRT; 109 AA.
AC Q81235;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ban2-3;
RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;
RT "The nucleotide sequence of the core region of HCV in Pakistan and
RT Bangladesh: The geographic characterization of HCV in South Asia.";
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D29652; BAA21027.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12235 MW; D5D1BBB10A7E3CAB CRC64;

Query Match
Best Local Similarity 91.7%; Score 143; DB 12; Length 109;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
5 PKPQPKRNTPRPDQVKFPGGGQIVG 32

```

```

RESULT 5
Q81233 PRELIMINARY; PRT; 109 AA.
AC Q81233;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ban2-1;
RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;

```

```

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D29647; BAA21022.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 12293 MW; 524E8D425A7E3CH2 CRC64;

Query Match
Best Local Similarity 91.7%; Score 143; DB 12; Length 109;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPDQVKFPGGGQIVG 28
   ||||| ||||| ||||| ||||| |||||
5 PKPQPKRNTPRPDQVKFPGGGQIVG 32

```

```

RESULT 5
Q81233 PRELIMINARY; PRT; 109 AA.
AC Q81233;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ban2-1;
RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;

```

RT "The nucleotide sequence of the core region of HCV in Pakistan and
 RT Bangladesh: The geographic characterization of HCV in South Asia.";
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL: D29650; BAA21025.1; -
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER 109 109

SQ SEQUENCE 109 AA; 12251 MW; D5DIA6060A7E3CAB CRC64;

Query Match 91.7%; Score 143; DB 12; Length 109;

Best Local Similarity 92.9%; Pred. No. 2.4e-12;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRPKNTPRRPQDVKFGGGQIVG 28

||||| ||||||| ||||||| ||||||| |||||||

5 PKPQRTKNTPRRPQNVKFGGGQIVG 32

RESULT 6

Q81231 ID Q81231 PRELIMINARY; PRT; 109 AA.

AC Q81231;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Ban1-2;

RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;

RT "The nucleotide sequence of the core region of HCV in Pakistan and

RT Bangladesh: The geographic characterization of HCV in South Asia.";

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL: D29648; BAA21023.1; -

DR InterPro: IPR002522; HCV_capsid.

DR Pfam: PF01543; HCV_capsid; 1.

DR Polyprotein.

FT NON_TER 109 109

SQ SEQUENCE 109 AA; 12261 MW; 461E99060A7E3CA9 CRC64;

Query Match 91.7%; Score 143; DB 12; Length 109;

Best Local Similarity 92.9%; Pred. No. 2.4e-12;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRPKNTPRRPQDVKFGGGQIVG 28

||||| ||||||| ||||||| ||||||| |||||||

5 PKPQRTKNTPRRPQNVKFGGGQIVG 32

RESULT 7

Q81232 ID Q81232 PRELIMINARY; PRT; 109 AA.

AC Q81232;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Ban1-3;

RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;

RT "The nucleotide sequence of the core region of HCV in Pakistan and

RT Bangladesh: The geographic characterization of HCV in South Asia.";

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL: D29649; BAA21024.1; -

DR InterPro: IPR002522; HCV_capsid.

DR Pfam: PF01543; HCV_capsid; 1.

DR Polyprotein.

FT NON_TER 109 109

SQ SEQUENCE 109 AA; 12261 MW; 461E99060A7E3CA9 CRC64;

Query Match 91.7%; Score 143; DB 12; Length 109;

Best Local Similarity 92.9%; Pred. No. 2.4e-12;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRPKNTPRRPQDVKFGGGQIVG 28

||||| ||||||| ||||||| ||||||| |||||||

5 PKPQRTKNTPRRPQNVKFGGGQIVG 32

RESULT 8

Q81234 ID Q81234 PRELIMINARY; PRT; 109 AA.

AC Q81234;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Core protein (Genome polyprotein) (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Ban2-2;

RA Ohno T., Mizokami M., Chstophor T., Ohba K., Suzuki K., Roger W.;

RT "The nucleotide sequence of the core region of HCV in Pakistan and

RT Bangladesh: The geographic characterization of HCV in South Asia.";

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL: D29651; BAA21026.1; -

DR InterPro: IPR002522; HCV_capsid.

DR Pfam: PF01543; HCV_capsid; 1.

DR Polyprotein.

FT NON_TER 109 109

SQ SEQUENCE 109 AA; 12251 MW; D5DIA6060A7E3CAB CRC64;

Query Match 91.7%; Score 143; DB 12; Length 109;

Best Local Similarity 92.9%; Pred. No. 2.4e-12;

Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRPKNTPRRPQDVKFGGGQIVG 28

||||| ||||||| ||||||| ||||||| |||||||

5 PKPQRTKNTPRRPQNVKFGGGQIVG 32

RESULT 9

Q81236 ID Q81236 PRELIMINARY; PRT; 109 AA.

AC Q81236;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

```

Db      2  PKPQRTKRNTPRPQNVKPPGGQIVG 29
||||| ||||| ||||| ||||| ||||| |||||
RESULT 11
ID Q68861 PRELIMINARY; PRT; 150 AA.
AC Q68861;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Capsid protein (Genome polyprotein) (Fragment).
GN HCV CORE.
OS Hepatitis C virus type 3b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
ON NCBI_TaxID=42791;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN-IND 308;
RX MEDLINE=96432288; PubMed=8835354;
RA Panigrahi A.K., Roca J., Acharya S.K., Jameel S., Panda S.K.;
RT "Genotype determination of hepatitis C virus from Northern India :
RL Identification of a new subtype."
RL J. Med. Virol. 48:191-198(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; X91299; CAA62673.1; -
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Polyprotein.
KW Polyprotein.
FT NON_TER 1
FT NON_TER 150
FT NON_TER 150
SQ SEQUENCE 150 AA; 16503 MW; 2C38B48C6BB9C16A CRC64;

Query Match 91.7%; Score 143; DB 12; Length 150;
Best Local Similarity 92.9%; Pred. No. 3.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0

QY 1 PKPQRPKRNTPRPQDVKPPGGQIVG 28
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 PKPQRTKRNTPRPQNVKPPGGQIVG 29

RESULT 12
ID Q68863 PRELIMINARY; PRT; 150 AA.
AC Q68863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Capsid protein (Genome polyprotein) (Fragment).
GN HCV CORE.
OS Hepatitis C virus type 3g.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
ON NCBI_TaxID=42792;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN-IND 1452;
RX MEDLINE=96432288; PubMed=8835354;
RA Panigrahi A.K., Roca J., Acharya S.K., Jameel S., Panda S.K.;
RT "Genotype determination of hepatitis C virus from Northern India :
RL Identification of a new subtype."
RL J. Med. Virol. 48:191-198(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).

```

```

DR EMBL: X91306; CAA62680.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 1 1
FT NON_TER 150 150
SQ SEQUENCE 150 AA; 16799 MW; BC2932134026E5FD CRC64;

Query Match          91.7%; Score 143; DB 12; Length 150;
Best Local Similarity 92.9%; Pred. No. 3.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRPQDVKFGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
DB 2 PKPQPKRNTPRRPQDVKFGGGQIVG 29

--SULT 13
1865
AC Q68865; PRELIMINARY; PRT; 150 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Capsid protein (Genome polyprotein) (Fragment).
GN HCV CORE.
OS Hepatitis C virus type 3g.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=42792;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IND 1751;
RX MEDLINE=96432288; PubMed=8835354;
RA Panigrahi A.K., Roca J., Acharya S.K., Jameel S., Panda S.K.;
RT "Genotype determination of hepatitis C virus from Northern India :
RT Identification of a new subtype.";
RL J. Med. Virol. 48:191-198(1996).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: X91423; CAA62759.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 1 1
FT NON_TER 150 150
SQ SEQUENCE 150 AA; 16728 MW; BC2D761707FCE1B9 CRC64;

Query Match          91.7%; Score 143; DB 12; Length 150;
Best Local Similarity 92.9%; Pred. No. 3.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRPQDVKFGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
DB 2 PKPQPKRNTPRRPQDVKFGGGQIVG 29

RESULT 14
Q8V7S2
ID Q8V7S2; PRELIMINARY; PRT; 191 AA.
AC Q8V7S2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.

```

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic
RT lesions.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AB062202; BAB83396.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191 191
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20740 MW; 64151CA83695EF34 CRC64;

Query Match          91.7%; Score 143; DB 12; Length 191;
Best Local Similarity 92.9%; Pred. No. 4.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRPQDVKFGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
DB 5 PKPQPKRNTNRRPQDVKFGGGQIVG 32

RESULT 15
Q8V7T0
ID Q8V7T0; PRELIMINARY; PRT; 191 AA.
AC Q8V7T0;
DT 01-WAR-2002 (TrEMBLrel. 20, Created)
DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic
RT lesions.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AB062194; BAB83388.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191 191
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20740 MW; 64151CA83695EF34 CRC64;

Query Match          91.7%; Score 143; DB 12; Length 191;
Best Local Similarity 92.9%; Pred. No. 4.4e-12;
Matches 26; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRPQDVKFGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
DB 5 PKPQPKRNTNRRPQDVKFGGGQIVG 32

```

Fri Aug 8 09:20:13 2003

Search completed: August 7, 2003, 11:19:00
Job time : 26.6364 secs

us-09-491-146a-27.rspt

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-27

Perfect score: 156

Sequence: 1 PKQRQPKRNTPRPQDKPFGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

al number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	28	3	US-08-921-887-27
2	139	89.1	28	3	US-08-921-887-24
3	139	89.1	74	3	US-08-836-075A-104
4	138	88.5	28	3	US-08-921-887-26
5	136	87.2	191	2	US-08-290-665A-175
6	136	87.2	191	5	PCT-US95-10398-175
7	135	86.5	34	3	US-08-380-160-6
8	135	86.5	43	4	US-09-020-846-36
9	135	86.5	44	3	US-08-380-160-2
10	135	86.5	44	4	US-09-389-756-1
11	135	86.5	45	3	US-08-380-160-1
12	135	86.5	61	1	US-07-946-054-9
13	135	86.5	61	1	US-08-083-947-23
14	135	86.5	61	1	US-08-530-550-3
15	135	86.5	61	1	US-08-262-037-26
16	135	86.5	61	5	PCT-US93-08638-9
17	135	86.5	61	5	PCT-US94-07088-23
18	135	86.5	61	5	PCT-US95-13660-3
19	135	86.5	74	3	US-08-836-075A-10
20	135	86.5	74	4	US-08-635-886C-198
21	135	86.5	100	3	US-08-635-886C-232
22	135	86.5	108	3	US-08-836-075A-14
23	135	86.5	115	1	US-08-324-977-8
24	135	86.5	115	2	US-08-384-616-8
25	135	86.5	115	2	US-08-904-686A-8
26	135	86.5	115	3	US-09-315-850-8
27	135	86.5	123	2	US-08-501-195-2

28 135 86.5 137 3 US-08-836-075A-46
29 135 86.5 138 3 US-08-836-075A-60
30 135 86.5 140 2 US-08-501-195-1
31 135 86.5 154 3 US-08-854-531-2
32 135 86.5 154 5 PCT-US95-13552-2
33 135 86.5 190 1 US-07-681-701-16
34 135 86.5 190 3 US-08-078-271B-1
35 135 86.5 191 2 US-08-290-665A-155
36 135 86.5 191 2 US-08-290-665A-156
37 135 86.5 191 2 US-08-290-665A-157
38 135 86.5 191 2 US-08-290-665A-158
39 135 86.5 191 2 US-08-290-665A-159
40 135 86.5 191 2 US-08-290-665A-160
41 135 86.5 191 2 US-08-290-665A-161
42 135 86.5 191 2 US-08-290-665A-163
43 135 86.5 191 2 US-08-290-665A-164
44 135 86.5 191 2 US-08-290-665A-165
45 135 86.5 191 2 US-08-290-665A-166

ALIGNMENTS

RESULT 1
US-08-921-887-27
; Sequence 27, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SA
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-27

Query Match 100.0%; Score 156; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 4.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRQDVKFGGQIVG 28
 Db 1 PKPQPKRNTPRRQDVKFGGQIVG 28

RESULT 2

US-08-921-887-24

; Sequence 24, Application US/08921887

; Patent No. 6030771

; GENERAL INFORMATION:

; APPLICANT: KHUDYAKOV, YURI E.

; APPLICANT: FIELDS, HOWARD A.

; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION

; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JONES & ASKEW, LLP

; STREET: 191 Peachtree Street, N.W., 37th Floor

; CITY: Atlanta

; STATE: GA

; COUNTRY: USA

; ZIP: 30303-1769

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/921,887

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: WARREN, WILLIAM L.

; REGISTRATION NUMBER: 36,714

; REFERENCE/DOCKET NUMBER: 03063-0380

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 404-818-3700

; TELEFAX: 404-818-3799

; INFORMATION FOR SEQ ID NO: 24:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: internal

; ORIGINAL SOURCE:

; ORGANISM: Hepatitis virus

US-08-921-887-24

Query Match

Best Local Similarity 89.1%; Score 139; DB 3; Length 28;

Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRQDVKFGGQIVG 28

Db 1 PKPQPKRNTPRRQDVKFGGQIVG 28

RESULT 3

US-08-836-075A-104

; Sequence 104, Application US/08836075A

; Patent No. 6180768

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: STUYVER, LIEVEN

; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC

; TITLE OF INVENTION: AGENTS

; NUMBER OF SEQUENCES: 207

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE & DURKEE

; STREET: P.O. BOX 4433

; CITY: HOUSTON

; STATE: TEXAS

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word 6.0 / ASCII text output

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/836.075A

; FILING DATE: 21 Apr 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP95/04155

; FILING DATE: 23 Oct 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 94870166.9

; FILING DATE: 21 Oct 1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 95870076.7

; FILING DATE: 28 Jun 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: KAMMERER, PATRICIA A.

; REGISTRATION NUMBER: 29,775

; REFERENCE/DOCKET NUMBER: INNS:004

; INFORMATION FOR SEQ ID NO: 104:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 74 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-836-075A-104

Query Match 89.1%; Score 139; DB 3; Length 74;

Best Local Similarity 92.9%; Pred. No. 2.1e-11;

Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRQDVKFGGQIVG 28

Db 5 PKPQPKRNTPRRQDVKFGGQIVG 32

RESULT 4

US-08-921-887-26

; Sequence 26, Application US/08921887

; Patent No. 6030771

; GENERAL INFORMATION:

; APPLICANT: KHUDYAKOV, YURI E.

; APPLICANT: FIELDS, HOWARD A.

; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION

; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JONES & ASKEW, LLP

; STREET: 191 Peachtree Street, N.W., 37th Floor

; CITY: Atlanta

; STATE: GA

; COUNTRY: USA

; ZIP: 30303-1769

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/921,887

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: WARREN, WILLIAM L.

REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 03063-0380
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis virus
US-08-921-887-26

Query Match 88.5%; Score 138; DB 3; Length 28;
Best Local Similarity 89.3%; Pred. No. 1e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRQDVKFGGGQIVG 28
|||||: ||| ||||| ||||| |||||
Db 1 PKPQPKRNTPRRQDVKFGGGQIVG 28

RESULT 5
US-08-290-665A-175
Sequence 175, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens

INDIVIDUAL ISOLATE: P8
US-08-290-665A-175
Query Match 87.2%; Score 136; DB 2; Length 191;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQPKRNTPRRQDVKFGGGQIVG 28
|||||: ||| ||||| ||||| |||||
Db 5 PKPQPKRNTPRRQDVKFGGGQIVG 32
RESULT 6
PCT-US95-10398-175
Sequence 175, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: P8
PCT-US95-10398-175
Query Match 87.2%; Score 136; DB 5; Length 191;
Best Local Similarity 89.3%; Pred. No. 1.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQPKRNTPRRQDVKFGGGQIVG 28
|||||: ||| ||||| ||||| |||||
Db 5 PKPQPKRNTPRRQDVKFGGGQIVG 32


```
RESULT 7
US-08-380-160-6
; Sequence 6, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,160
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 28682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-6400
; TELEFAX: (703)836-2787
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
US-08-380-160-6
Query Match 86.5%; Score 135; DB 3; Length 34;
Best Local Similarity 89.3%; Pred. No. 3.1e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRPKNTPRPDVKFPGGGQIVG 28
Db 4 PKPQRTKNTNRRPDVKFPGGGQIVG 31
;
RESULT 8
US-09-020-846-36
; Sequence 36, Application US/09020846
; Patent No. 6322965
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, Kenjiro
; APPLICANT: KASHIWAKUMA, Tomiko
; APPLICANT: CHIBA, Yukie
; APPLICANT: YAGI, Shintaro
; APPLICANT: HASEGAWA, Akira
; TITLE OF INVENTION: CHIMERA ANTIGEN PEPTIDE
;
US-08-380-160-6
Query Match 86.5%; Score 135; DB 3; Length 34;
Best Local Similarity 89.3%; Pred. No. 3.1e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRPKNTPRPDVKFPGGGQIVG 28
Db 4 PKPQRTKNTNRRPDVKFPGGGQIVG 31
;
RESULT 9
US-08-380-160-2
; Sequence 2, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,160
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-027015
; FILING DATE: 10-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-024045
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 053466/0225
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-020-846-36
Query Match 86.5%; Score 135; DB 4; Length 43;
Best Local Similarity 89.3%; Pred. No. 3.9e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRPKNTPRPDVKFPGGGQIVG 28
Db 5 PKPQRTKNTNRRPDVKFPGGGQIVG 32
;
RESULT 9
US-08-380-160-2
; Sequence 2, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,160
```

```
; FILING DATE: 530
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 28682
; TELEPHONE: (703)836-6400
; TELEFAX: (703)836-2787
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
; STRAIN: H77
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..44
; OTHER INFORMATION: /note= "N-terminal sequence of the
; OTHER INFORMATION: protein of the nucleocapside or CORE protein of
; OTHER INFORMATION: the human hepatitis C virus"
;
US-08-380-160-2

Query Match 86.5%; Score 135; DB 3; Length 44;
Best Local Similarity 89.3%; Pred. No. 4e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRPODVKFGGGQIVG 28
    |||||: ||||| ||||| ||||| |||||
Db 4 PKPQKTKRNTNRRPODVKFGGGQIVG 31

RESULT 10
US-09-389-756-1
; Sequence 1, Application US/09389756
; Patent No. 6576240
; GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/09/389,756
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: PCT/FR98/00442
; EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; Patent No. 6576240
; TITLE: of Hepatitis Virus
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
```

```
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-09-389-756-1

Query Match 86.5%; Score 135; DB 4; Length 44;
Best Local Similarity 89.3%; Pred. No. 4e-11;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRRPODVKFGGGQIVG 28
    |||||: ||||| ||||| ||||| |||||
Db 4 PKPQKTKRNTNRRPODVKFGGGQIVG 31

RESULT 11
US-08-380-160-1
; Sequence 1, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY;
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,160
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 28682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-6400
; TELEFAX: (703)836-2787
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
; STRAIN: H77
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..45
; OTHER INFORMATION: /note= "N-terminal sequence of the
; OTHER INFORMATION: protein of the nucleocapside or CORE protein of
; OTHER INFORMATION: the human hepatitis C virus"
;
US-08-380-160-1
```

Query Match 86.5%; Score 135; DB 3; Length 45;
 Best Local Similarity 89.3%; Pred. No. 4.1e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPQDVKFGGGQIVG 28
 Db 5 PKPQKTKRNTNRPRQDVKFGGGQIVG 32

RESULT 12

US-07-946-054-9
 ; Sequence 9, Application US/07946054
 ; Patent No. 5582968
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Chang Yi
 ; APPLICANT: Hosein, Barbara H
 ; TITLE OF INVENTION: No. 5582968el Branched Hybrid and Cluster
 ; TITLE OF INVENTION: Peptides Effective in Diagnosing and Detecting No. 5582968-A,
 ; TITLE OF INVENTION: No. 5582968-B Hepatitis
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: United Biomedical Inc.
 ; STREET: 25 Davids Dr.
 ; CITY: Hauppauge
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 11788

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/946.054
 FILING DATE: 15-SEP-1992
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, M. Lisa
 REGISTRATION NUMBER: 34,045
 REFERENCE/DOCKET NUMBER: 2000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 516-273-2828
 TELEFAX: 516-273-1717
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide

07-946-054-9

Query Match 86.5%; Score 135; DB 1; Length 61;
 Best Local Similarity 89.3%; Pred. No. 5.7e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPQDVKFGGGQIVG 28
 Db 4 PKPQKTKRNTNRPRQDVKFGGGQIVG 31

RESULT 13

US-08-083-947-23
 ; Sequence 23, Application US/08083947
 ; Patent No. 5639594
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Chang Yi
 ; APPLICANT: Hosein, Barbara
 ; TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
 ; TITLE OF INVENTION: In Diagnosing and Detecting No. 5639594-A, No. 5639594-B Hepat
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: M. Lisa Wilson

STREET: 25 Davids Drive
 CITY: Hauppauge
 STATE: NY
 COUNTRY: USA
 ZIP: 11788
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/083.947
 FILING DATE: 19930628
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 946,054
 FILING DATE: 15-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, M. Lisa
 REGISTRATION NUMBER: 34045
 REFERENCE/DOCKET NUMBER: 2000Z
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516)273-2828
 TELEFAX: (516)273-1717
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide

US-08-083-947-23

Query Match 86.5%; Score 135; DB 1; Length 61;
 Best Local Similarity 89.3%; Pred. No. 5.7e-11;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQPKRNTPRPQDVKFGGGQIVG 28
 Db 4 PKPQKTKRNTNRPRQDVKFGGGQIVG 31

RESULT 14

US-08-530-550-3
 ; Sequence 3, Application US/08530550
 ; Patent No. 5736321
 ; GENERAL INFORMATION:
 ; APPLICANT: Hosein, Barbara
 ; APPLICANT: Wang, Chang Yi
 ; TITLE OF INVENTION: Peptides Effective for Diagnosis and
 ; TITLE OF INVENTION: Detection of Hepatitis c Infection
 ; NUMBER OF SEQUENCES: 51
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: M. Lisa Wilson
 ; STREET: 25 Davids Drive
 ; CITY: Hauppauge
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 11788

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/530,550
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, M. Lisa
 REGISTRATION NUMBER: 34,045
 REFERENCE/DOCKET NUMBER: 2000Z
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516)273-2828

TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-530-550-3

Query Match 86.5%; Score 135; DB 1; Length 61;

Best Local Similarity 89.3%; Pred. No. 5,7e-11;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRRPDVKFPGGGQIVG 28
|||||: ||||| ||||||||| |||||
Db 4 PKPQRTKRNTRRRPDVKFPGGGQIVG 31

RESULT 15

--08-262-037-26

Sequence 26, Application US/08262037

Patent No. 5747239

GENERAL INFORMATION:

APPLICANT: Chang Yi Wang and Barbara Hosein

TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR

TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV

TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES

NUMBER OF SEQUENCES: 136

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVE.

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/262,037

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/719,819

FILING DATE: 24-June-1991

APPLICATION NUMBER: 07/667,275

FILING DATE: 11-Mar-1991

APPLICATION NUMBER: 07/651,735

FILING DATE: 07-Feb-1991

APPLICATION NUMBER: 07/558,799

FILING DATE: 26-July-1990

APPLICATION NUMBER: 07/510,153

FILING DATE: 16-April-1990

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4043 US3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 61 amino acids

TYPE: Amino acid

STRANDEDNESS:

TOPOLOGY: Unknown

US-08-262-037-26

Query Match

Best Local Similarity

86.5%; Score 135; DB 1; Length 61;

89.3%; Pred. No. 5,7e-11;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPQRPKRNTPRRPDVKFPGGGQIVG 28
|||||: ||||| ||||||||| |||||
Db 4 PKPQRTKRNTRRRPDVKFPGGGQIVG 31

Search completed: August 7, 2003, 11:23:51

Job time : 11.6364 secs

SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-09-851-138-104

Query Match 89.1%; Score 139; DB 10; Length 74;
Best Local Similarity 92.9%; Pred. No. 1.8e-10;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRPQDVKFGGGQIVG 28
DB 5 PFPQRTKRNTRRPQDVKFGGGQIVG 32

RESULT 2

US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
GENERAL INFORMATION:
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANCOIS
; APPLICANT: DALBON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
; TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
; TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
; TITLE OF INVENTION: TREATING AN HCV INFECTION
; FILE REFERENCE: 103959
; CURRENT APPLICATION NUMBER: US/10/367,677
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US/09/389,756
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; PUBLICATION INFORMATION:
; AUTHORS: Ogata, N. et al.
; TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 3392-3396
; DATE: 1991
; RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match 86.5%; Score 135; DB 15; Length 44;
Best Local Similarity 89.3%; Pred. No. 3.4e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRPQDVKFGGGQIVG 28
DB 4 PKPQRTKRNTRRPQDVKFGGGQIVG 31

RESULT 3

US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US20020183508A1
GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P. O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-851-138-10

Query Match 86.5%; Score 135; DB 10; Length 74;
Best Local Similarity 89.3%; Pred. No. 5.7e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRPQDVKFGGGQIVG 28
DB 5 PKPQRTKRNTRRPQDVKFGGGQIVG 32

RESULT 4

US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US20020090607A1
GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 14114.034902
; CURRENT APPLICATION NUMBER: US/09/758,308
; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 86.5%; Score 135; DB 9; Length 91;
Best Local Similarity 89.3%; Pred. No. 7.1e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQRPKRNTPRPQDVKFGGGQIVG 28
DB 5 PKPQRTKRNTRRPQDVKFGGGQIVG 32

RESULT 5
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756.875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8
Query Match 86.5%; Score 135; DB 10; Length 97;
Best Local Similarity 89.3%; Pred. No. 7.6e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 1 PKPQKPKRNTPRRPQDVKFPGGQIVG 28
5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32
RESULT 6
US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: STD nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; TITLE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-921-397-77
Query Match 86.5%; Score 135; DB 10; Length 103;
Best Local Similarity 89.3%; Pred. No. 8.1e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQKPKRNTPRRPQDVKFPGGQIVG 28
Db 18 PKPQKTKRNTNRRPQDVKFPGGQIVG 45
RESULT 7
US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851.138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-851-138-14
Query Match 86.5%; Score 135; DB 10; Length 108;
Best Local Similarity 89.3%; Pred. No. 8.5e-10;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQKPKRNTPRRPQDVKFPGGQIVG 28
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32
RESULT 8
US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS

Query Match	86.5%	Score 135;	DB 10;	Length 137;
Best Local Similarity	89.3%	Pred. No. 1.1e-09;		

MEDIUM TYPE: Floppy disk


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-152

```

```

Query Match      86.5%; Score 135; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 1.3e-09;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db      1 PKPQRPKRNTPRRPQDVKFPGGGQIVG 28
        |||||: ||||| ||||| ||||| |||||
        5 PKPQRTKRNTNRRPQDVKFPGGGQIVG 32

```

```

RESULT 12
US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-152

```

```

Query Match      86.5%; Score 135; DB 11; Length 166;
Best Local Similarity 89.3%; Pred. No. 1.3e-09;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db      1 PKPQRPKRNTPRRPQDVKFPGGGQIVG 28
        |||||: ||||| ||||| ||||| |||||
        5 PKPQRTKRNTNRRPQDVKFPGGGQIVG 32

```

```

RESULT 13
US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-42

```

```

Query Match      86.5%; Score 135; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 1.3e-09;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 PKQRPKRNTPRRPQDVKFPGGGQIVG 28
        |||||: ||||| ||||| ||||| |||||
        5 PKPQRTKRNTNRRPQDVKFPGGGQIVG 32

```

```

RESULT 14
US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US200300308274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-44

```

```

Query Match      86.5%; Score 135; DB 11; Length 169;
Best Local Similarity 89.3%; Pred. No. 1.3e-09;
Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 PKQRPKRNTPRRPQDVKFPGGGQIVG 28
        |||||: ||||| ||||| ||||| |||||
        5 PKPQRTKRNTNRRPQDVKFPGGGQIVG 32

```

```

RESULT 15
US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-42

Query Match      86.5%; Score 135; DB 11; Length 169;
%st Local Similarity 89.3%; Pred. No. 1.3e-09;
atches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 PKPQKPKRNTPRRPQDVKFGGGQIVG 28
      |||||: ||||| ||||| ||||| |||||
Db      5 PKPQKTKRNTNRRPQDVKFGGGQIVG 32

Search completed: August 7, 2003, 12:01:13
Job time : 14.3636 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:37 ; Search time 38.5455 Seconds
(without alignments)
115.301 Million cell updates/sec

Title: US-09-491-146A-25

Perfect score: 154
Sequence: 1 PPKPQKTKNTYRRPODVKFPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
al number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	100.0	28	20	AA06675 NC mosaic protein
2	154	100.0	3010	15	AA06675 Blood transmiscibl
3	154	100.0	3010	23	AA06675 HCV-S1 full-length
4	146	94.8	28	20	AA06675 NC mosaic protein
5	146	94.8	189	23	AA06675 HCV type 3 capsid
6	146	94.8	191	17	AA06675 Hepatitis C virus
7	146	94.8	191	17	AA06675 Hepatitis C virus
8	146	94.8	191	17	AA06675 Hepatitis C virus
9	146	94.8	319	17	AA06675 Hepatitis C virus

10	145	94.2	36	16	AA06675	CN14 fragment of H
11	145	94.2	38	14	AA06675	HCV capsid peptide
12	145	94.2	38	14	AA06675	HCV capsid peptide
13	145	94.2	38	14	AA06675	HCV capsid peptide
14	145	94.2	38	15	AA06675	Non-A, non-B hepat
15	145	94.2	40	16	AA06675	Hepatitis C virus
16	145	94.2	40	16	AA06675	Hepatitis C virus
17	145	94.2	43	19	AA06675	Hepatitis C virus
18	145	94.2	44	19	AA06675	Hepatitis C virus
19	145	94.2	44	20	AA06675	Hepatitis C virus
20	145	94.2	44	21	AA06675	Human hepatitis C
21	145	94.2	45	21	AA06675	Human hepatitis C
22	145	94.2	50	16	AA06675	Hepatitis C virus
23	145	94.2	55	13	AA06675	HCV core-envelope
24	145	94.2	55	13	AA06675	HCV core-envelope
25	145	94.2	55	13	AA06675	HCV core-envelope
26	145	94.2	55	13	AA06675	HCV core-envelope
27	145	94.2	55	13	AA06675	HCV core-envelope
28	145	94.2	57	13	AA06675	Non-A, Non-B hepat
29	145	94.2	61	16	AA06675	Peptide VIIIE base
30	145	94.2	61	16	AA06675	Anti-HCV antibody
31	145	94.2	61	17	AA06675	Prototype peptide
32	145	94.2	61	18	AA06675	HCV core protein p
33	145	94.2	66	12	AA06675	PT-NANB viral stru
34	145	94.2	74	17	AA06675	Hepatitis C virus
35	145	94.2	78	13	AA06675	Antigen pHCal01.
36	145	94.2	78	13	AA06675	Antigen pHCal01.
37	145	94.2	79	14	AA06675	HCV fragment 1 / 1
38	145	94.2	79	14	AA06675	HCV fragment 2 / 1
39	145	94.2	80	15	AA06675	Branched peptide H
40	145	94.2	82	13	AA06675	Non-A, Non-B hepat
41	145	94.2	82	13	AA06675	Non-A, Non-B hepat
42	145	94.2	90	16	AA06675	Hepatitis C virus
43	145	94.2	91	22	AA06675	Antigenic epitope
44	145	94.2	97	14	AA06675	HCV core protein N
45	145	94.2	97	16	AA06675	Hepatitis C virus

ALIGNMENTS

RESULT 1
AA06675

ID AA06675 standard; Protein; 28 AA.

XX AA06675;

DT 17-JUN-1999 (first entry)

XX NC mosaic protein amino acid fragment C.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
restriction endonuclease assisted ligation; vaccination.

OS Hepatitis C virus.

XX WO9910506-A1.

PD 04-MAR-1999.

PF 21-AUG-1998; 98WO-US17385.

PR 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1999-204671/17.

XX New mosaic protein, comprising a plurality of homologous antigenic
peptides from different genotypes of a species - useful for
detecting hepatitis infection in an individual

XX

PS Claim 5; Fig 9; 66pp; English.

XX CC The invention relates to a mosaic protein, comprising a plurality of
 CC homologous antigenic peptides from different genotypes of a species. The
 CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
 CC synthesizing an artificial gene that encodes the mosaic protein is also
 CC provided. The method is designated restriction endonuclease assisted
 CC ligation (REAL). The mosaic protein and the artificial mosaic protein
 CC are useful for detecting a hepatitis infection in an individual. The
 CC mosaic gene and protein is also useful for vaccination against
 CC infection, especially hepatitis C. The method of synthesizing the
 CC artificial gene and the resulting mosaic protein improve the sensitivity,
 CC spectrum of immunoreactivity, and antigen specificity of enzyme
 CC immunoassays. This provides improved detection of hepatitis C virus.
 CC Sequences AAY06673-683 represent amino acid sequence of each monomer
 CC comprising the NC mosaic protein.

XX SQ Sequence 28 AA;

Very Match 100.0%; Score 154; DB 20; Length 28;
 st Local Similarity 100.0%; Pred. NO. 1.2e-14;
 atches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVAFPGGGQIVG 28

Db 1 PKPQKTKRNTYRRPDQVAFPGGGQIVG 28

RESULT 2

AAR53417

ID AAR53417 standard; Protein; 3010 AA.

XX AC AAR53417;

XX DT 17-JAN-1995 (first entry)

XX DE Blood transmissible NANBH virus protein.

XX KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
 KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
 KW C100 antibody; HCV RNA; NS5 region.

XX OS Non-A, non-B hepatitis virus.

XX FH Key Location/Qualifiers

FT Misc-difference 222

FT /label= His, Arg

FT Misc-difference 226

FT /label= Cys, Arg

FT Misc-difference 246

FT /label= Leu, Phe

FT Misc-difference 263

FT /label= Asp, Asn

FT Misc-difference 291

FT /label= Phe, Ser

FT Misc-difference 311

FT /label= Gly, Asp

FT Misc-difference 398

FT /label= Ser, Arg, Gly

FT Misc-difference 400

FT /label= Thr, Ala

FT Misc-difference 405

FT /label= Gln, Pro, Leu

FT Misc-difference 410

FT /label= Lys, Arg

FT Misc-difference 418

FT /label= Gly, Asp

FT Misc-difference 430

FT /label= Asn, Asp

FT Misc-difference 438

FT /label= Phe, Leu

FT Misc-difference 478

FT Misc-difference 759 /label= Arg, Lys
 FT /label= Leu, Val
 FT Misc-difference 1017 /label= Ser, Asn
 FT Misc-difference 1036 /label= Thr, Ala
 FT Misc-difference 1056 /label= Glu, Asp
 FT Misc-difference 1201 /label= Met, Thr
 FT Misc-difference 1205 /label= Met, Ile
 FT Misc-difference 1255 /label= Asn, Tyr
 FT Misc-difference 1263 /label= Gly, Asp
 FT Misc-difference 1455 /label= Asn, Asp
 FT Misc-difference 1828 /label= Ala, Thr
 FT Misc-difference 1895 /label= Gly, Arg
 FT Misc-difference 1896 /label= Gly, Ile
 FT Misc-difference 2143 /label= Glu, Val
 FT Misc-difference 2144 /label= Asp, Glu
 FT Misc-difference 2462 /label= Cys, Arg
 FT Misc-difference 2486 /label= Val, Met
 FT Misc-difference 2488 /label= Lys, Gln
 FT Misc-difference 2844 /label= Leu, Met
 FT Misc-difference 2862 /label= Leu, Gln
 FT Misc-difference 2917 /label= Arg, Leu
 FT Misc-difference 2968 /label= Ser, Gly
 FT Misc-difference 2989 /label= Cys, Arg
 FT Misc-difference 2990 /label= Tyr, Cys

JP06105690-A.

19-APR-1994.

10-MAR-1992; 92JP-0051885.

10-MAR-1992; 92JP-0051885.

(KAEN/) KAENNO K.

WPI; 1994-163130/20.

N-PSDB; AAQ63499.

Blood-transmissible non-A non-B hepatitis virus DNA - used for
 detection of hepatitis virus

Claim 1; Page 8-20; 22pp; Japanese.

This sequence is encoded by the genome of a blood transmissible non-A,
 non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the
 primers given in AAQ63500-35. The amplified fragments are used in the
 detection of hepatitis virus. The target DNA was isolated from serum
 of chronically infected NANBH patients who were C100 antibody-positive
 and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR
 were performed on cDNA and the total human NANBH DNA was constructed

CC from 23 clones.

XX Sequence 3010 AA;

Query Match 100.0%; Score 154; DB 15; Length 3010;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKPQRTKRNTRYRRPDVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRYRRPDVKFPGGGQIVG 32

RESULT 3

AAE20477
ID AAE20477 standard; Protein; 3010 AA.

XX AAE20477;

XX 01-JUL-2002 (first entry)

XX HCV-SI full-length polyprotein.

XX Nucleic acid construct; expression cassette; non-coding region; NCR;
KW untranslated region; UTR; anti-viral drug; drug resistance;
KW HCV-SI; Hepatitis C virus.

XX Hepatitis C virus.

XX WO200208447-A2.

XX 31-JAN-2002.

XX 20-JUL-2001; 2001WO-IL00669.

XX 24-JUL-2000; 2000US-220248P.

XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.
PA (EHRICH) EHRICH G.

XX Tan YH, Lim SP, Lim SG, Hong WJ;

XX WPI; 2002-280605/32.

XX N-PSDB; AAD33038.

XX Novel nucleic acid construct useful for detecting the presence of RNA virus, comprises an expression cassette and a promoter operably linked to expression cassette for minus strand RNA transcription of the cassette

XX Example 1; Page 70-81; 81pp; English.

XX The invention relates to nucleic acid construct which comprises an expression cassette including a first polynucleotide region including a 5' non-coding region (NCR) sequence of an RNA virus and at least an N-terminal portion of a coding sequence of RNA virus, a second polynucleotide region including a 3' untranslated region (UTR) sequence of the RNA virus and at least a C-terminal portion of a coding sequence of the RNA virus and a third polynucleotide region encoding a reporter molecule, flanked by first and second polynucleotide regions; and a promoter sequence being operatively linked to expression cassette in a manner so as to enable a transcription of a minus strand RNA molecule from the expression cassette. Nucleic acid construct of the invention is useful for detecting the presence of an RNA virus in a cell. It is also useful for screening anti-viral drugs and determining drug resistance of an RNA virus. The present sequence is Hepatitis C virus (HCV) isolate HCV-SI full-length polyprotein.

XX Sequence 3010 AA;

Query Match 100.0%; Score 154; DB 23; Length 3010;

Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKPQRTKRNTRYRRPDVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRYRRPDVKFPGGGQIVG 32

RESULT 4

AAO06673
ID AAO06673 standard; Protein; 28 AA.

XX AAO06673;

XX 17-JUN-1999 (first entry)

XX NC mosaic protein amino acid fragment A.

XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
KW restriction endonuclease assisted ligation; vaccination.

XX Hepatitis C virus.

XX WO9910506-A1.

XX 04-MAR-1999.

XX 21-AUG-1998; 98WO-US17385.

XX 25-AUG-1997; 97US-0921887.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1999-204671/17.

XX New mosaic protein, comprising a plurality of homologous antigenic peptides from different genotypes of a species - useful for detecting hepatitis infection in an individual

XX Claim 5; Fig 9; 66pp; English.

XX The invention relates to a mosaic protein, comprising a plurality of homologous antigenic peptides from different genotypes of a species. The antigenic peptides are from nucleocapsid (NC) proteins. A method for synthesizing an artificial gene that encodes the mosaic protein is also provided. The method is designated restriction endonuclease assisted ligation (REAL). The mosaic protein and the artificial mosaic protein are useful for detecting a hepatitis infection in an individual. The mosaic gene and protein is also useful for vaccination against infection, especially hepatitis C. The method of synthesizing the artificial gene and the resulting mosaic protein improve the sensitivity, spectrum of immunoreactivity, and antigen specificity of enzyme immunoassays. This provides improved detection of hepatitis C virus. Sequences AAO06673-683 represent amino acid sequence of each monomer comprising the NC mosaic protein.

XX Sequence 28 AA;

Query Match 94.8%; Score 146; DB 20; Length 28;

Best Local Similarity 96.4%; Pred. No. 1.6e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQRTKRNTRYRRPDVKFPGGGQIVG 28
Db 1 PKPQRTKRNTRYRRPDVKFPGGGQIVG 28

RESULT 5

AAO71258
ID AAO71258 standard; protein; 189 AA.

XX AAO71258;

XX AAO71258;

```

DT 18-NOV-2002 (first entry)
XX HCV type 3 capsid protein fragment.
XX Capsid protein; attenuated vaccine; virucide; antiinflammatory;
KW hepatotropic; yellow fever; Japanese encephalitis; dengue;
KW classical swine fever; bovine viral diarrhoea; hepatitis C.
XX Hepatitis c virus.
XX WO200266621-A1.
XX 29-AUG-2002.
XX 11-FEB-2002; 2002WO-AT00046.
XX 21-FEB-2001; 2001AT-0000272.
XX (HEINZ) HEINZ F X.
XX (MANDL) MANDL C.
XX Heinz FX, Mandl C;
XX WPI; 2002-667064/71.
XX Attenuated flavivirus live vaccine, useful for protection against e.g.
XX yellow fever, comprises virus with attenuating deletion of amino acids
XX from the capsid protein.
XX Disclosure; Fig 2; 30pp; German.
XX This invention describes a novel attenuated flavivirus live vaccine
XX comprising a flavivirus mutant that has a deletion of at least 4
XX consecutive amino acids from the capsid protein, provided that the
XX C-terminal hydrophobic region is not affected by the deletion. The
XX vaccine of the invention has virucide, antiinflammatory and hepatotropic
XX activity. The attenuated vaccine, and similar nucleic acid vaccines that
XX encode the mutated capsid protein, are useful for protection against a
XX wide range of flavivirus diseases, e.g. yellow fever, Japanese
XX encephalitis, dengue, classical swine fever, bovine viral diarrhoea and
XX hepatitis C. The specified deletion: (i) produces a reliably attenuated
XX virus that does not revert to virulence; (ii) is exactly defined and does
XX not effect immune responses to important proteins; and (iii) can not
XX generate a non-natural virus by recombination. The mutant viruses
XX eliminate the need to produce large amounts of infectious/virulent
XX viruses, and can be produced with less expense. The protective response
XX to flavivirus lasts significantly longer than that to killed vaccines.
XX This sequence represents a fragment of the capsid protein from Hepatitis
XX c virus (HCV) type 1 described in the disclosure of the invention.
XX
XX Sequence 189 AA;
XX
XX Query Match 94.8%; Score 146; DB 23; Length 189;
XX Best Local Similarity 96.4%; Pred. No. 1.1e-12;
XX Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 PKPQKTKRNTYRRPQDVKFPFGGQIVG 28
XX | | | | | | | | | | | | | | | | | |
XX DB 4 PKPQKTKRNTYRRPQDVKFPFGGQIVG 31
XX
XX RESULT 6
XX AAR92968
XX ID AAR92968 standard; Protein; 191 AA.
XX
XX AC AAR92968;
XX
XX DT 02-OCT-1996 (first entry)
XX Hepatitis C virus isolate HK10 core protein.
XX
XX DE Hepatitis C virus isolate HK10 core protein.
XX
XX KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX
XX RESULT 7
XX AAR92969
XX ID AAR92969 standard; Protein; 191 AA.
XX
XX AC AAR92969;
XX
XX DT 02-OCT-1996 (first entry)
XX Hepatitis C virus isolate S52 core protein.
XX
XX DE Hepatitis C virus isolate S52 core protein.
XX
XX KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX
XX OS Hepatitis C virus.
XX
XX PN WO9605315-A2.
XX
XX PD 22-FEB-1996.
XX
XX PF 15-AUG-1995; 95WO-US10398.
XX
XX PR 15-AUG-1994; 94US-0290665.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI (USSH ) US SEC DEPT HEALTH.
XX
XX PI Bukh J, Miller RH, Purcell RH;
XX
XX WPI; 1996-139709/14.
XX
XX DR N-PSDB; NAT16642.
XX
XX DNA and amino acid sequence of HCV envelope 1 and core proteins -
XX used to determine HCV genotype and as vaccines against HCV infection
XX Claim 4; Page 207; 340pp; English.
XX
XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX isolates. Isolated cDNA sequences are used for the prodn. of primers
XX useful for detecting the presence of HCV in a sample, the primers
XX are also useful for HCV genotyping. Proteins encoded by the cDNAs
XX can be used in vaccines for immunising against HCV infection. The
XX proteins may also be used to detect antibodies against HCV in serum,
XX saliva, lymphocytes or other mononuclear cells. The antibodies may
XX be used in the prevention of HCV infection.
XX
XX SQ Sequence 191 AA;
XX
XX Query Match 94.8%; Score 146; DB 17; Length 191;
XX Best Local Similarity 96.4%; Pred. No. 1.1e-12;
XX Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 PKPQKTKRNTYRRPQDVKFPFGGQIVG 28
XX | | | | | | | | | | | | | | | | | |
XX DB 5 PKPQKTKRNTYRRPQDVKFPFGGQIVG 32
XX
XX RESULT 7
XX AAR92969
XX ID AAR92969 standard; Protein; 191 AA.
XX
XX AC AAR92969;
XX
XX DT 02-OCT-1996 (first entry)
XX Hepatitis C virus isolate S52 core protein.
XX
XX DE Hepatitis C virus isolate S52 core protein.
XX
XX KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX hepatitis.
XX
XX OS Hepatitis C virus.
XX
XX PN WO9605315-A2.
XX
XX PD 22-FEB-1996.
XX
XX PF 15-AUG-1995; 95WO-US10398.
XX
XX PR 15-AUG-1994; 94US-0290665.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI (USSH ) US SEC DEPT HEALTH.
XX
XX PI Bukh J, Miller RH, Purcell RH;
XX
XX WPI; 1996-139709/14.
XX

```

DR N-PSDB; AAT16643.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

XX

XX Claim 4; Page 208; 340pp; English.

XX

CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers

CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX

XX Sequence 191 AA;

SQ

Query Match 94.8%; Score 146; DB 17; Length 191;

Best Local Similarity 96.4%; Pred. No. 1.1e-12;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTYRRPQDVKPPGGQIVG 28

||||||| ||||||| ||||||| |||||||

DB 5 PKPQRTKRTYRRPQDVKPPGGQIVG 32

||||||| ||||||| ||||||| |||||||

RESULT 8

AAR92971

ID AAR92971 standard; Protein; 191 AA.

XX

AC AAR92971;

XX

XX 02-OCT-1996 (first entry)

DT

XX

XX Hepatitis C virus isolate DK12 core protein.

DE

XX HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

KW

XX Hepatitis C virus.

OS

XX WO9605315-A2.

PN

XX

PD 22-FEB-1996.

XX

XX 15-AUG-1995; 95WO-US10398.

PF

XX

... 15-AUG-1994; 94US-0290665.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (USSH) US SEC DEPT HEALTH.

XX

PI Bukh J, Miller RH, Purcell RH;

XX

XX WPI; 1996-139709/14.

DR N-PSDB; AAT16645.

XX

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

XX

XX Claim 4; Page 209-210; 340pp; English.

XX

CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers

CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The

CC proteins may also be used to detect antibodies against HCV in serum,

CC saliva, lymphocytes or other mononuclear cells. The antibodies may

CC be used in the prevention of HCV infection.

XX

XX Sequence 191 AA;

SQ

Query Match 94.8%; Score 146; DB 17; Length 191;

Best Local Similarity 96.4%; Pred. No. 1.1e-12;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTYRRPQDVKPPGGQIVG 28

||||||| ||||||| ||||||| |||||||

DB 5 PKPQRTKRTYRRPQDVKPPGGQIVG 32

||||||| ||||||| ||||||| |||||||

RESULT 9

AAR96547

ID AAR96547 standard; peptide; 319 AA.

XX

AC AAR96547;

XX

XX 10-MAR-1997 (first entry)

DT

XX

XX Hepatitis C virus types 7c(8a) isolates VN4 amino acids 1-317.

DE

XX

XX Hepatitis C virus; subtype; polymerase chain reaction; amplification;

KW PCR; primer; probe; antibody; infection.

KW

XX

OS Hepatitis C virus.

XX

XX Key Location/Qualifiers

FT Misc-difference 144

FT /label= Met, Leu

FT Misc-difference 144..149

FT /label= Val, Ala, Glu, Gly

FT Misc-difference 156

FT /label= Met, Thr, Lys, Arg

FT Misc-difference 157

FT /label= Val, Ala, Asp, Gly

FT Misc-difference 161

FT /label= Gly

FT /note= "amino acid in this position is designated X in

FT the specification, but codon usage shows that the

FT only possible amino acid at this pos. is Gly"

FT

FT Misc-difference 167

FT /label= Ser, Arg

FT Misc-difference 167

FT /label= Val, Ala, Glu, Gly

FT Misc-difference 171

FT /label= Gly

FT /note= "amino acid in this position is designated X in

FT the specification, but codon usage shows that the

FT only possible amino acid at this pos. is Gly"

FT

FT Misc-difference 172

FT /label= Cys, Arg, Ser, Gly

FT Misc-difference 174

FT /label= Phe, Leu, Ile, Val

FT Misc-difference 177

FT /label= Phe, Leu

FT Misc-difference 232

FT /label= Met, Val

FT Misc-difference 233

FT /label= Asn, Asp

XX

XX WO9613590-A2.

PN

XX

XX 09-MAY-1996.

PD

XX

XX 23-OCT-1995; 95WO-EP04155.

PF

XX

XX 28-JUN-1995; 95EP-0870076.

PR

XX

XX 21-OCT-1994; 94EP-0870166.

PR

XX (INNO-) INNOGENETICS NV.

PA

XX Maertens G, Stuyver L;

PI

XX

XX WPI; 1996-251460/25.

DR

DR N-PSDB; AAT27958.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:types
 PT - used to develop probes and primers for new sub:types and vaccines
 PT to prevent and treat infection
 XX
 XX Claim 25; Fig 3; 150pp; English.
 XX
 CC The sequences AAR96526-R96578 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
 CC genome. This sequence represents amino acids 1-317 from the HCV types 7c
 CC and 8a isolates VN4.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 XX
 XX Sequence 319 AA;
 CC
 CC Query Match 94.8%; Score 146; DB 17; Length 319;
 CC Best Local Similarity 96.4%; Pred. No. 1.9e-12;
 CC Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 PKPQRTKRTNRYRRPDVKFPGGGQIVG 28
 CC DB 5 PKPQRTKRTNRYRRPDVKFPGGGQIVG 32
 CC
 CC RESULT 10
 CC AAW06487
 CC ID AAW06487 standard; Protein; 36 AA.
 CC AC AAW06487;
 CC XX
 CC XX 31-JAN-1997 (first entry)
 CC DT
 CC DE CN14 fragment of HCV core region.
 CC XX
 CC XX CN14; CP14; core region; hepatitis C virus; HCV; detection;
 CC KW antibody.
 CC
 CC Synthetic.
 CC XX
 CC PN JP06327482-A.
 CC XX
 CC PD 29-NOV-1994.
 CC XX
 CC PF 21-MAY-1993; 93JP-0156026.
 CC XX
 CC PR 21-MAY-1993; 93JP-0156026.
 CC XX
 CC PA (IMMO) IMMUNO JAPAN KK.
 CC XX
 CC XX WPI; 1995-047903/07.
 CC DR N-PSDB; AAT45055.
 CC XX
 CC XX Detection of hepatitis C virus - using oligopeptide fragment of HCV
 CC PT core region
 CC PT
 CC XX
 CC PS Claim 1; Page 6; 7pp; Japanese.
 CC XX
 CC XX This sequence is encoded by the oligonucleotide, CN14, and represents
 CC CC the peptide fragment CP14. CP14 is a fragment of the core region
 CC of hepatitis C virus (HCV). CP14 may be used in the detection of

CC HCV infection and to raise antibodies against it.
 XX
 XX Sequence 36 AA;
 CC
 CC Query Match 94.2%; Score 145; DB 16; Length 36;
 CC Best Local Similarity 96.4%; Pred. No. 2.9e-13;
 CC Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 PKPQRTKRTNRYRRPDVKFPGGGQIVG 28
 CC DB 1 PKPQRTKRTNRYRRPDVKFPGGGQIVG 28
 CC
 CC RESULT 11
 CC AAR30687
 CC ID AAR30687 standard; peptide; 38 AA.
 CC AC AAR30687;
 CC XX
 CC XX 25-MAR-2003 (updated)
 CC DT 11-MAY-1993 (first entry)
 CC XX
 CC DE HCV capsid peptide No. 23.
 CC XX
 CC KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
 CC KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;
 CC KW non-A, non-B hepatitis; competitive; inhibition assay.
 CC XX
 CC OS Hepatitis C virus.
 CC PN WO9222571-A1.
 CC XX
 CC PD 23-DEC-1992.
 CC XX
 CC PF 29-APR-1992; 92WO-US03635.
 CC XX
 CC PR 13-JUN-1991; 91US-0714471.
 CC PR 20-JUN-1991; 91US-0718052.
 CC XX
 CC PA (BAXT) BAXTER DIAGNOSTICS INC.
 CC XX
 CC PI Jolley ME, Leahy DC, Todd JA;
 CC XX
 CC DR WPI; 1993-018073/02.
 CC XX
 CC PT Synthetic peptide epitope with sequence encoded by hepatitis-C
 CC PT virus - for immunoassay for antigens for diagnosis of non-A,
 CC PT non-B hepatitis
 CC XX
 CC PS Disclosure; Fig 1F; 66pp; English.
 CC XX
 CC CC The sequences given in AAR30655-89 represent fragments of the
 CC Hepatitis C virus (HCV) amino acid sequence. They represent the
 CC beginning of the HCV open reading frame to amino acid 38 and
 CC encompass the "common" sequence. These peptides are contained in
 CC the capsid protein of the virus and themselves contain epitope
 CC groups. These peptides can be used in immunoassays for HCV
 CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
 CC in competitive inhibition assay for detecting HCV specific
 CC antibodies.
 CC CC (Updated on 25-MAR-2003 to correct PN field.)
 CC XX
 CC XX Sequence 38 AA;
 CC
 CC Query Match 94.2%; Score 145; DB 14; Length 38;
 CC Best Local Similarity 96.4%; Pred. No. 3e-13;
 CC Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 PKPQRTKRTNRYRRPDVKFPGGGQIVG 28
 CC DB 5 PKPQRTKRTNRYRRPDVKFPGGGQIVG 32

RESULT 12

AAR30688
ID AAR30688 standard; peptide; 38 AA.

XX AC AAR30688;

XX DT 25-MAR-2003 (updated)

XX ET 11-MAY-1993 (first entry)

XX DE HCV capsid peptide No. 24.

XX KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;
KW non-A, non-B hepatitis; competitive; inhibition assay.

XX OS Hepatitis C virus.

XX PN WO9222571-A1.

XX PT 23-DEC-1992.

XX PS 29-APR-1992; 92WO-US03635.

XX PR 13-JUN-1991; 91US-0714471.

XX PT 20-JUN-1991; 91US-0718052.

XX PA (BAXT) BAXTER DIAGNOSTICS INC.

XX PI Jolley ME, Leahy DC, Todd JA;

XX DR WPI; 1993-018073/02.

XX PT Synthetic peptide epitope with sequence encoded by hepatitis-C
PT virus - for immunoassay for antigens for diagnosis of non-A,
PT non-B hepatitis

XX PS Disclosure; Fig 1F; 66pp; English.

XX CC The sequences given in AAR30665-89 represent fragments of the
CC Hepatitis C virus (HCV) amino acid sequence. They represent the
CC beginning of the HCV open reading frame to amino acid 38 and
CC encompass the "common" sequence. These peptides are contained in
CC the capsid protein of the virus and themselves contain epitope
CC groups. These peptides can be used in immunoassays for HCV
CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
CC in competitive inhibition assay for detecting HCV specific
CC antibodies.
CC (Updated on 25-MAR-2003 to correct PN field.)

CC

Sequence 38 AA;

Query Match 94.2%; Score 145; DB 14; Length 38;

Best Local Similarity 96.4%; Pred. No. 3e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTYRRPDQVRFPGGGQIVG 28

DB 5 PKPQRTKNTNRRPDQVRFPGGGQIVG 32

RESULT 13

AAR30689
ID AAR30689 standard; peptide; 38 AA.

XX AC AAR30689;

XX DT 25-MAR-2003 (updated)

XX ET 11-MAY-1993 (first entry)

XX DE HCV capsid peptide No. 25.

XX KW Hepatitis C virus; HCV; open reading frame; "common" sequence;
KW capsid protein; epitope; immunoassay; antibody; diagnosis; NANBH;

KW non-A, non-B hepatitis; competitive; inhibition assay.

XX OS Hepatitis C virus.

XX PN WO9222571-A1.

XX PT 23-DEC-1992.

XX PF 29-APR-1992; 92WO-US03635.

XX PR 13-JUN-1991; 91US-0714471.

XX PT 20-JUN-1991; 91US-0718052.

XX PA (BAXT) BAXTER DIAGNOSTICS INC.

XX PI Jolley ME, Leahy DC, Todd JA;

XX DR WPI; 1993-018073/02.

XX PT Synthetic peptide epitope with sequence encoded by hepatitis-C
PT virus - for immunoassay for antigens for diagnosis of non-A,
PT non-B hepatitis

XX PS Disclosure; Fig 1F; 66pp; English.

XX CC The sequences given in AAR30665-89 represent fragments of the
CC Hepatitis C virus (HCV) amino acid sequence. They represent the
CC beginning of the HCV open reading frame to amino acid 38 and
CC encompass the "common" sequence. These peptides are contained in
CC the capsid protein of the virus and themselves contain epitope
CC groups. These peptides can be used in immunoassays for HCV
CC antibodies in the diagnosis of non-A, non-B hepatitis (NANBH), and
CC in competitive inhibition assay for detecting HCV specific
CC antibodies.
CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 38 AA;

Query Match 94.2%; Score 145; DB 14; Length 38;

Best Local Similarity 96.4%; Pred. No. 3e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTYRRPDQVRFPGGGQIVG 28

DB 5 PKPQRTKNTNRRPDQVRFPGGGQIVG 32

RESULT 14

AAR54065
ID AAR54065 standard; Protein; 38 AA.

XX AC AAR54065;

XX DT 14-FEB-1995 (first entry)

XX DE Non-A, non-B hepatitis virus core1 region fragment.

XX KW Non-A, non-B hepatitis virus; NANBH; hepatitis C virus; HCV;
KW core; ENV; NS1; NS2; NS3; antigen; detection.

XX OS Hepatitis C virus.

XX PN JP06141870-A.

XX PD 24-MAY-1994.

XX PF 12-MAR-1992; 92JP-0088140.

XX PR 12-MAR-1992; 92JP-0088140.

XX PA (SANW) SANWA KAGAKU KENKYUSHO CO.

XX PA (TOFU) TONEN CORP.

XX PA (TORR-) ZH TOKYO RINSHO IGAKU SOGO KENKYUSHO.

Search completed: August 7, 2003, 11:14:06
Job time : 38.6364 secs

```

XX  WPI: 1994-205028/25.
DR  N-PSDB; AAQ64067.
XX
PT  DNA coding a Non-A, non-B hepatitis virus antigen - useful for
PT  detecting HCV within serum
XX
PS  Disclosure; Page 11; 22pp; Japanese.
XX
CC  Hepatitis C virus #4 and #6 genes were isolated (AAQ64068-69).
CC  Both genes contain the core, ENV, NS1, NS2 and NS3 regions.
CC  A core region fragment is given in AAQ64067.
XX
SQ  Sequence 38 AA;

Query Match      94.2%; Score 145; DB 15; Length 38;
Best Local Similarity 96.4%; Pred. No. 3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qv  1 PKPQRTKRNTYRRPQDVKEFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
    5 PKPQRTKRNTYRRPQDVKEFGGQIVG 32

RESULT 15
AAR84559
ID  AAR84559 standard; peptide; 40 AA.
XX
AC  AAR84559;
XX
DT  06-JUN-1996 (first entry)
XX
DE  Hepatitis C virus core antigenic peptide.
XX
KW  HCV; non-A, non-B hepatitis virus; antigen; immunoassay;
KW  interferon treatment; monitoring; antibody titre; detection.
XX
OS  Hepatitis C virus.
XX
PN  JP07260792-A.
XX
PD  13-OCT-1995.
XX
PF  16-MAR-1994; 94JP-0082160.
XX
PR  16-MAR-1994; 94JP-0082160.
XX
PA  (ARIM/) ARIMA T.
PA  (TORA ) TORAY IND INC.
XX
WPI; 1995-386063/50.
XX
PT  Determination of the antibody titre against hepatitis C virus
PT  antigen - using dilute solution and HCV antigenic peptide(s)
XX
PS  Claim 1; Page 10; 11pp; Japanese.
XX
CC  The present sequence is that of an antigen used in a novel assay for
CC  determining anti-HCV antibody titre. A sample is diluted to an
CC  absorbance of not more than 2.0 and the peptide antigen is added.
CC  The method is useful for monitoring the effect of interferon
CC  treatment on anti-HCV antibody titres.
XX
SQ  Sequence 40 AA;

Query Match      94.2%; Score 145; DB 16; Length 40;
Best Local Similarity 96.4%; Pred. No. 3.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  1 PKPQRTKRNTYRRPQDVKEFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
    4 PKPQRTKRNTYRRPQDVKEFGGQIVG 31
Db

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 : Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146a-25

Perfect score: 154

Sequence: 1 FKPQKTRNTYRRPDVKFPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
al number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	3010	1 A45573	genome polyprotein
2	146	94.8	114	2 S41359	genome polyprotein
3	146	94.8	114	2 S41358	genome polyprotein
4	146	94.8	124	2 S41360	genome polyprotein
5	146	94.8	411	2 PC2061	genome polyprotein
6	146	94.8	492	2 S41288	genome polyprotein
7	145	94.2	88	2 S21336	genome polyprotein
8	145	94.2	108	2 S41353	genome polyprotein
9	145	94.2	108	2 S41355	genome polyprotein
10	145	94.2	108	2 S41357	genome polyprotein
11	145	94.2	108	2 S41356	genome polyprotein
12	145	94.2	108	2 S41348	genome polyprotein
13	145	94.2	112	2 S41371	genome polyprotein
14	145	94.2	112	2 S41341	genome polyprotein
15	145	94.2	114	2 S41370	genome polyprotein
16	145	94.2	114	2 S41369	genome polyprotein
17	145	94.2	114	2 S41368	genome polyprotein
18	145	94.2	115	2 S41342	genome polyprotein
19	145	94.2	115	2 S41344	genome polyprotein
20	145	94.2	115	2 S41350	genome polyprotein
21	145	94.2	115	2 S41354	genome polyprotein
22	145	94.2	115	2 S41345	genome polyprotein
23	145	94.2	115	2 S41347	genome polyprotein
24	145	94.2	115	2 S41343	genome polyprotein
25	145	94.2	118	2 S41346	genome polyprotein
26	145	94.2	369	2 S21471	genome polyprotein
27	145	94.2	441	2 S21207	genome polyprotein
28	145	94.2	513	2 PC1284	genome polyprotein
29	145	94.2	520	2 JQ1925	polyprotein - hepa

30 145 94.2 523 2 JQ1926
31 145 94.2 550 2 JH0711
32 145 94.2 782 2 S19876
33 145 94.2 782 2 S18031
34 145 94.2 782 2 S18032
35 145 94.2 787 2 JN0677
36 145 94.2 874 2 JQ0883
37 145 94.2 874 2 JQ0881
38 145 94.2 876 2 PC2219
39 145 94.2 3010 1 GNVVTC
40 145 94.2 3010 1 GNVVCJ
41 145 94.2 3010 1 S18030
42 145 94.2 3011 1 GNVVCH
43 145 94.2 3011 1 S40770
44 145 94.2 3014 1 JC5620
45 145 94.2 3033 1 JQ1303

ALIGNMENTS

RESULT 1

A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (non)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hiji
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese car
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g22161
A:Experimental source: HCV-JT
A:Note: Sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:730-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1230-1237/Product: hepacivirin #status predicted <NS3>
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 100.0%; Score 154; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 2.8e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPQKTRNTYRRPDVKFPGGGQIVG 28

DB 5 FKPQKTRNTYRRPDVKFPGGGQIVG 32

RESULT 2

S41359
genome polyprotein - hepatitis C virus (genotype 3, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41359
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41359
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:Z29462; NID:q443886; PIDN:CAA82600.1; PID:q443887
A:Experimental source: genotype 3, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match 94.8%; Score 146; DB 2; Length 114;
Best Local Similarity 96.4%; Pred. No. 1.5e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 3
genome polyprotein - hepatitis C virus (genotype 3, N1) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41358
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41358
A:Molecule type: genomic RNA
A:Residues: 1-114 <VAN>
A:Cross-references: EMBL:Z29461; NID:q443884; PIDN:CAA82599.1; PID:q443885
A:Experimental source: genotype 3, N1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <MAT>

Query Match 94.8%; Score 146; DB 2; Length 114;
Best Local Similarity 96.4%; Pred. No. 1.5e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 4
genome polyprotein - hepatitis C virus (genotype 3, N3) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N3
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41360
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41360
A:Molecule type: genomic RNA
A:Residues: 1-124 <VAN>
A:Cross-references: EMBL:Z29463; NID:q443888; PIDN:CAA82601.1; PID:q443889
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-124/Product: core protein #status predicted <MAT>

Query Match 94.8%; Score 146; DB 2; Length 124;

Best Local Similarity 96.4%; Pred. No. 1.6e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 5
genome polyprotein N2 - hepatitis C virus
N:Contains: envelope protein E1; nonstructural protein E2/NS1
C:Species: hepatitis C virus
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
C:Accession: PC2061
R:Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A:Title: Identification of the third major genotype of hepatitis C virus in France.
A:Reference number: PC2060; MUID:94197744; PMID:8147893
A:Accession: PC2061
A:Molecule type: mRNA
A:Residues: 1-411 <LIJ>
A:Cross-references: GB:L12355; NID:g410169; PIDN:AAA20155.1; PID:g410170
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein
F:192-383/Product: envelope protein E1 #status predicted <SPE>
F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NPE>
F:196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 94.8%; Score 146; DB 2; Length 411;
Best Local Similarity 96.4%; Pred. No. 5.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 6
genome polyprotein - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein; NS1 protein
C:Species: hepatitis C virus
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41288
R:Seelig, R.
submitted to the EMBL Data Library, December 1993
A:Reference number: S41288
A:Accession: S41288
A:Molecule type: genomic RNA
A:Residues: 1-492 <SEE>
C:Cross-references: EMBL:X76918
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; envelope protein; nonstructural protein;
F:1-191/Product: core protein #status predicted <COR>
F:192-372/Product: envelope protein #status predicted <ENV>
F:373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Query Match 94.8%; Score 146; DB 2; Length 492;
Best Local Similarity 96.4%; Pred. No. 6.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 7
genome polyprotein S29 (core protein region) - hepatitis C virus (fragment)
C:Species: hepatitis C virus
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 17-Nov-2000
C:Accession: S21336

R:Sato, A.
 submitted to the EMBL Data Library, April 1992
 A:Description: A sensitive serodiagnosis of hepatitis C virus infection with two cloned
 A:Reference number: S21336
 A:Accession: S21336
 A:Molecule type: genomic RNA
 A:Residues: 1-88 <SAT>
 A:Cross-references: EMBL:X65548; NID:g59492; PIDN:CAA46517.1; PID:g59493
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: polyprotein

Query Match 94.2%; Score 145; DB 2; Length 88;
 Best Local Similarity 96.4%; Pred. No. 1.9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTYRRPQDVFPFGGGQIVG 28
 |||||
 Db 11 PKPQRTKNTYRRPQDVFPFGGGQIVG 38

JULT 8

..1353
 genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N2
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41353
 R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41353
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:Z29456
 A:Experimental source: genotype 2, N2
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <NAT>

Query Match 94.2%; Score 145; DB 2; Length 108;
 Best Local Similarity 96.4%; Pred. No. 1.9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTYRRPQDVFPFGGGQIVG 28
 |||||
 ~ 5 PKPQRTKNTYRRPQDVFPFGGGQIVG 32

RESULT 9

S41355
 genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N4
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41355
 R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41355
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:Z29458
 A:Experimental source: genotype 2, N4
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <NAT>

Query Match 94.2%; Score 145; DB 2; Length 108;
 Best Local Similarity 96.4%; Pred. No. 1.9e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQRTKNTYRRPQDVFPFGGGQIVG 28
 |||||
 Db 5 PKPQRTKNTYRRPQDVFPFGGGQIVG 32

RESULT 10

S41357
 genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N6
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41357
 R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41357
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:Z29460
 A:Experimental source: genotype 2, N6
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <NAT>

Query Match 94.2%; Score 145; DB 2; Length 108;
 Best Local Similarity 96.4%; Pred. No. 1.9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTYRRPQDVFPFGGGQIVG 28
 |||||
 Db 5 PKPQRTKNTYRRPQDVFPFGGGQIVG 32

RESULT 11

S41356
 genome polyprotein - hepatitis C virus (genotype 2, N5) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 2, N5
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41356
 R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41356
 A:Molecule type: genomic RNA
 A:Residues: 1-108 <VAN>
 A:Cross-references: EMBL:Z29459
 A:Experimental source: genotype 2, N5
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: capsid protein; core protein; polyprotein
 F:1-108/Product: core protein #status predicted <NAT>

Query Match 94.2%; Score 145; DB 2; Length 108;
 Best Local Similarity 96.4%; Pred. No. 1.9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKNTYRRPQDVFPFGGGQIVG 28
 |||||
 Db 5 PKPQRTKNTYRRPQDVFPFGGGQIVG 32

RESULT 12

S41348
 genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N6

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41348
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41348

A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:Z29451
A:Experimental source: genotype 1, N6
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <Mat>

Query Match 94.2%; Score 145; DB 2; Length 108;
Best Local Similarity 96.4%; Pred. No. 1.9e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QV 1 PKPQRTKRNTRYRRPQDVKFPGGQIVG 28
|||||
5 PKPQRTKRNTRYRRPQDVKFPGGQIVG 32

RESULT 13

S41371 genome polyprotein - hepatitis C virus (genotype 5, N5) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N5
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41371

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341

A:Accession: S41371
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
A:Cross-references: EMBL:Z29474
A:Experimental source: genotype 5, N5
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <Mat>

Query Match 94.2%; Score 145; DB 2; Length 112;
Best Local Similarity 96.4%; Pred. No. 2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 PKPQRTKRNTRYRRPQDVKFPGGQIVG 28
|||||
5 PKPQRTKRNTRYRRPQDVKFPGGQIVG 32

RESULT 14

S41341

genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41341
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341

A:Accession: S41341
A:Molecule type: genomic RNA
A:Residues: 1-112 <VAN>
A:Cross-references: EMBL:Z29444; NID:g443850; PIDN:CAA82582.1; PID:g443851
A:Experimental source: genotype 1, N1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein

F:1-112/Product: core protein #status predicted <Mat>

Query Match 94.2%; Score 145; DB 2; Length 112;
Best Local Similarity 96.4%; Pred. No. 2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QV 1 PKPQRTKRNTRYRRPQDVKFPGGQIVG 28
|||||
5 PKPQRTKRNTRYRRPQDVKFPGGQIVG 32

RESULT 15

S41370

genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)

N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41370
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341

A:Accession: S41370

A:Molecule type: genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: EMBL:Z29473; NID:g443908; PIDN:CAA82611.1; PID:g443909
A:Experimental source: genotype 5, N4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-114/Product: core protein #status predicted <Mat>

Query Match 94.2%; Score 145; DB 2; Length 114;
Best Local Similarity 96.4%; Pred. No. 2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QV 1 PKPQRTKRNTRYRRPQDVKFPGGQIVG 28
|||||
5 PKPQRTKRNTRYRRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 11:21:47
Job time : 9.54545 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 4.90909 Seconds

(without alignments)
268.226 Million cell updates/sec

Title: US-09-491-146a-25

Perfect score: 154

Sequence: 1 PKQRKTRNTYRRPQDVKFPGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
tal number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	3010	1	POLG_HCVJT
2	145	94.2	513	1	POLG_HCVJ2
3	145	94.2	520	1	POLG_HCVH4
4	145	94.2	520	1	POLG_HCVHK
5	145	94.2	737	1	POLG_HCVJ5
6	145	94.2	737	1	POLG_HCVJ7
7	145	94.2	3010	1	POLG_HCVBK
8	145	94.2	3010	1	POLG_HCVJA
9	145	94.2	3011	1	POLG_HCVH
10	145	94.2	3033	1	POLG_HCVJ6
11	145	94.2	3033	1	POLG_HCVJ8
12	138	89.6	3010	1	POLG_HCVTW
13	137	89.0	794	1	POLG_HCV1
14	60	39.0	794	1	FURI_HUMAN
15	55	35.7	793	1	FURI_MOUSE
16	54	35.1	797	1	FURI_BOVIN
17	51	33.1	548	1	SYN_BRUMA
18	51	33.1	793	1	FURI_RAT
19	51	33.1	1398	1	NGO3_MOUSE
20	50	32.5	282	1	RK4_TOBAC
21	49	31.8	194	1	RS7_FUGRU
22	49	31.8	194	1	RS7_HUMAN
23	49	31.8	512	1	IE63_HSV2H
24	48	31.2	922	1	B3AT_CHICK
25	47.5	30.8	753	1	YBV2_YEAST
26	47	30.5	137	1	RL16_SPICI
27	47	30.5	485	1	ZDH1_MOUSE
28	46.5	30.2	303	1	YB1_XENLA
29	46	29.9	190	1	RS7_MANSE
30	46	29.9	194	1	RS7_XENLA
31	46	29.9	293	1	RK4_SPIOL
32	46	29.9	509	1	YF08_MYCPN
33	46	29.9	556	1	PDPK_HUMAN

34 46 29.9 559 1 PDPK_MOUSE
35 46 29.9 559 1 PDPK_RAT
36 46 29.9 676 1 RNR_CHLPN
37 46 29.9 1311 1 FMR2_HUMAN
38 45.5 29.5 3988 1 POLG_BVDVN
39 45 29.2 108 1 UL39_DROME
40 45 29.2 195 1 RS7_SCHPO
41 45 29.2 317 1 SOX2_HUMAN
42 45 29.2 319 1 SOX2_MOUSE
43 45 29.2 320 1 SOX2_SHEEP
44 45 29.2 454 1 NCAP_CVM3
45 45 29.2 454 1 NCAP_CVMA5

ALIGNMENTS

RESULT 1
POLG_HCVJT STANDARD; PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2;
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).
OS Hepatitis C virus (Isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.; PubMed-1318627;
RX MEDLINE-9295714; PubMed-1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D11168; BAA01943.1; -
CC PIR: A45573; A45573.
CC PDB: 1A1Q; 25-MAR-98.
CC PDB: 1JXP; 14-JAN-98.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC InterPro: IPR001410; DEAD.
CC

InterPro: IPR002522; HCV_capsid.
InterPro: IPR002521; HCV_core.
InterPro: IPR002519; HCV_env.
InterPro: IPR002531; HCV_NSI.
InterPro: IPR002518; HCV_NS1.
InterPro: IPR002518; HCV_NS2.
InterPro: IPR004109; HCV_NS3.
InterPro: IPR000745; HCV_NS4a.
InterPro: IPR001490; HCV_NS4b.
InterPro: IPR002868; HCV_NS5a.
InterPro: IPR002166; HCV_RdRP.
InterPro: IPR007095; RNA_pol_DS_PS.
InterPro: IPR007094; RNA_pol_PSVir.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NSI; 1.
Pfam: PF01538; HCV_NS2; 1.
Pfam: PF02907; HCV_NS3; 1.
Pfam: PF01006; HCV_NS4a; 1.
Pfam: PF01001; HCV_NS4b; 1.
Pfam: PF01306; HCV_NS5a; 1.
Pfam: PF00271; helicase_C; 1.
Pfam: PF00998; Viral_RdRP; 1.
SMART: PD186062; HCV_NSI; 1.
SMART: SM00487; DEXdc; 1.
Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN 3010 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
FT SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
Query Match 100.0%; Score 154; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 1.3e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PKPQRTKNTYRRPDQVFKPGGQIVG 28

Db 5 PKPQRTKNTYRRPDQVFKPGGQIVG 32
|||||
RESULT 2
POLG_HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11111;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC EMBL: D10074; BAA00968.1;
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR ProDom: PD186062; HCV_NSI; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >513
FT CHAIN >513
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 513 513
FT NON_TER 513


```

FT  NON_TER      520
SQ  SEQUENCE      520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match          94.2%; Score 145; DB 1; Length 520;
Best Local Similarity 96.4%; Pred. No. 4.1e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 PKPQRTKRNTYRRPDQVKFPGGQIVG 28
Db  |||||  |||||  |||||  |||||  |||||
    5 PKPQRTKRNTYRRPDQVKFPGGQIVG 32

RESULT 4
POLG_HCVHK
ID  POLG_HCVHK      STANDARD;      PRT;      520 AA.
AC  Q01403;
DT  01-JUL-1993 (Rel. 26, Created)
DT  01-JUL-1993 (Rel. 26, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE  Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE  (GP68) (GP70) (NS1) (Fragment)].
OS  Hepatitis C virus (isolate HCV-HP) (HCV).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
OX  NCBI_TaxID=31644;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE=93019030; PubMed=1383400;
RA  Abe K., Inchauspe G., Fujisawa K.;
RT  "Genomic characterization and mutation rate of hepatitis C virus
RT  isolated from a patient who contracted hepatitis during an epidemic
RT  of non-A, non-B hepatitis in Japan.";
RT  J. Gen. Virol. 73:2725-2729(1992).
CC  -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC  LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC  PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC  PROTEIN C AND MRNA.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announcement/)
CC  or send an email to license@lsb-sib.ch).
CC  -----
EMBL; D10687; BAA01529.1; --
PIR; JQ1925; JQ1925.
InterPro; IPR002522; HCV_Capsid.
InterPro; IPR002521; HCV_Core.
InterPro; IPR002519; HCV_env.
InterPro; IPR002531; HCV_NS1.
Pfam; PF01543; HCV_capsid; 1.
Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_env; 1.
Pfam; PF01560; HCV_NS1; 1.
ProDom; PD186062; HCV_NS1; 1.
Polyprotein; Glycoprotein; Coat protein; Envelope protein;
Transmembrane; Nonstructural protein.
INIT_MET 1 1
      REMOVED FROM CAPSID PROTEIN C BY THE
      CELLULAR AMINOPEPTIDASE.
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT  CARBOHYD      234 234
FT  CARBOHYD      305 305
FT  CARBOHYD      418 418
FT  CARBOHYD      418 418
FT  CHAIN          1 115
FT  CHAIN          116 191
FT  CHAIN          192 383
FT  CHAIN          384 >520
FT  CHAIN          347 369
FT  TRANSMEM      195 196
FT  CARBOHYD      209 209
FT  CARBOHYD      233 233
FT
```

```

FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56475 MW; 1D2BD0A6FF27349B CRC64;

Query Match 94.2%; Score 145; DB 1; Length 520;
Best Local Similarity 96.4%; Pred. No. 4.1e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
DB 5 PPKQKTKRNTNRPPQDVKFPFGGGQIVG 32

RESULT 5
POLG_HCVJ5 STANDARD; PRT; 737 AA.
ID POLG_HCVJ5
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
DE 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (fragment).
DE Hepatitis C virus (isolate HC-J5) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10075; BAA00969.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 734 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

```

TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 94.2%; Score 145; DB 1; Length 737;
Best Local Similarity 96.4%; Pred. No. 6.1e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
DB 5 PPKQKTKRNTNRPPQDVKFPFGGGQIVG 32

RESULT 6
POLG_HCVJ7 STANDARD; PRT; 737 AA.
ID POLG_HCVJ7
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DE 01-AUG-1992 (Rel. 23, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (fragment).
DE Hepatitis C virus (isolate HC-J7) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11114;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10075; BAA00969.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 734 733 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 >737
 FT TRANSHEM 347 369
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 299 299
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT NON_TER 737 737
 SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 94.2%; Score 145; DB 1; Length 737;
 Best Local Similarity 96.4%; Pred. No. 6.1e-14;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PRPQRTKNTYRPPDVFPPGGQIVG 28
 DB 5 PRPQRTKNTNRRPDVFPFGGQIVG 32

RESULT 7
 POLG_HCVBK STANDARD; PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11105;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91140698; PubMed-18477440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers".
 RL J. Virol. 65:1105-1113(1991).
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE-96235224; PubMed-8647104;
 RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetky L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits

phosphorylation mediated by cAMP-dependent protein kinase.";
 Eur. J. Biochem. 237:611-618(1996).
 [3]
 RT X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RN MEDLINE-97015088; PubMed-8861916;
 RX Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
 RA Moosaw E.W., Adachi T., Hostomsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site.";
 Cell 87:331-342(1996).
 [4]
 RT X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RN MEDLINE-98227846; PubMed-9568891;
 RX Yan Y., Li F., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 Protein Sci. 7:837-847(1998).
 CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

 DR EMBL; M58335; AAA72945.1; .
 DR PIR; A38465; GNVTC.
 DR PDB; 1A10; 25-MAR-98.
 DR PDB; 1JXP; 14-JAN-98.
 DR PDB; 1NS3; 08-APR-98.
 DR PDB; 1C2P; 15-NOV-00.
 DR PDB; 1CSJ; 08-NOV-99.
 DR PDB; 1GX5; 09-APR-02.
 DR PDB; 1GX6; 10-APR-02.
 DR PDB; 1QVU; 26-JUN-00.
 DR PDB; 8OHM; 20-APR-99.
 DR MEROPS; S29.001; .
 DR MEROPS; U39.001; .
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXOC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN E (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS3 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS4 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1039 1047
 FT STRAND 1050 1050
 FT STRAND 1059 1063
 FT STRAND 1068 1074
 FT TURN 1075 1076
 FT STRAND 1077 1081
 FT HELIX 1082 1085
 FT TURN 1086 1087
 FT STRAND 1090 1092
 FT TURN 1093 1094
 FT STRAND 1095 1097
 FT STRAND 1101 1103
 FT TURN 1104 1107
 FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT TURN 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186

FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;
 Query Match 94.2%; Score 145; DB 1; Length 3010;
 Best Local Similarity 96.4%; Pred. No. 2.9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
 DB 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32
 RESULT 8
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91089550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC {RNA}(N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90208; BAA14233.1; -;
 DR PIR: A39253; GNMVCJ.

DR HSP: P26663; LJP.
 DR MEROPS: S29.001; -.
 DR MEROPS: U39.001; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_Capsid.
 DR InterPro: IPR002521; HCV_Core.
 DR InterPro: IPR002519; HCV_Env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5b.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_Capsid; 1.
 DR Pfam: PF01542; HCV_Core; 1.
 DR Pfam: PF01539; HCV_Env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural
 DR INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 94.2%; Score 145; DB 1; Length 3010;
 Best Local Similarity 96.4%; Pred. No. 2.9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 PKPQRTKRTNRRPDQVKKPGGQIVG 28
 Db 5 PKPQRTKRTNRRPDQVKKPGGQIVG 32
 RESULT 9
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 OX NCBI_TaxID=11108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RA "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -|- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -|- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -|- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -|- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -|- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -|- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -|- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -|- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

P26660;
 01-AUG-1992 (Rel. 23, Created)
 01-AUG-1992 (Rel. 23, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 ON NCBI_TaxID=11113;
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.,
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D00944; BAA00792.1;
 CC PIR: J01303; J01303.
 CC HSP: P27958; IHEI.
 CC MEROPS: S29.001;
 CC MEROPS: U39.001;
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_Psvir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.

DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NP_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 FT SSQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Query Match 94.2%; Score 145; DB 1; Length 3033;
 Best Local Similarity 96.4%; Pred. No. 2,9e-13;
 Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 PKPQKTKRNTYRRPDQVKFPGGQIVG 28
 Db 5 PKPQKTKRNTYRRPDQVKFPGGQIVG 32
 RESULT 11
 POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 ID POLG_HCVJ8
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 ON NCBI_TaxID=11115;
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
 Machida A., Miyakawa Y., Mayumi M.,
 "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 and divergent regions.";
 J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D00944; BAA00792.1;
 CC PIR: J01303; J01303.
 CC HSP: P27958; IHEI.
 CC MEROPS: S29.001;
 CC MEROPS: U39.001;
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_Psvir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.

SEQUENCE FROM N.A.
 MEDLINE-92230232; PubMed-1314459;
 Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 Fukuda S., Tsuda F., Mishiro S.;
 *Full-length sequence of a hepatitis C virus genome having poor
 homology to reported isolates: comparative study of four distinct
 genotypes.;
 Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: D10988; BAA01761.1; -
 PIR: A40250; GNMVJ8.
 DR HSP: P27958; IHEI.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RORP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RGRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXdc; 1.
 DR PolyProtein: Glycoprotein; Transferase: RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033

CHAIN 3033
 TRANSMEM 347 369
 ACT_SITE 1087 1097
 ACT_SITE 1111 1111
 ACT_SITE 1169 1169
 NP_BIND 1234 1241
 SITE 1320 1323
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 233 233
 CARBOHYD 299 299
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 477 477
 CARBOHYD 534 534
 CARBOHYD 542 542
 CARBOHYD 558 558
 CARBOHYD 578 578
 CARBOHYD 627 627
 CARBOHYD 649 649
 CARBOHYD 1091 1091
 CARBOHYD 2038 2038
 CARBOHYD 2359 2359
 CARBOHYD 2811 2811
 SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 94.2%; Score 145; DB 1; Length 3033;

Best Local Similarity 96.4%; Pred. No. 2.9e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28

Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 12

POLG_HCVTW

ID POLG_HCVTW STANDARD; PRT; 3010 AA.

AC P29846;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Taiwan) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus

OX NCBI_taxid-31645;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-92230206; PubMed-1314449;

RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

RT "The Taiwanese hepatitis C virus genome: sequence determination and

RT mapping the 5' termini of viral genomic and antigenomic RNA.";

RL Virology 188:102-113(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC [RNA](N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

DR PIR: A39166; GNWVC3.
 DR PDB: 1AIV; 16-FEB-99.
 DR PDB: 1HEI; 25-NOV-98.
 DR MEROPS: S29.001; .
 DR MEROPS: U39.001; .
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR002490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 ne Pfam: PF01543; HCV_capsid; 1.
 Pfam: PF01542; HCV_core; 1.
 Pfam: PF01539; HCV_env; 1.
 Pfam: PF01560; HCV_NS1; 1.
 Pfam: PF01538; HCV_NS2; 1.
 Pfam: PF02907; HCV_NS3; 1.
 Pfam: PF01006; HCV_NS4a; 1.
 Pfam: PF01001; HCV_NS4b; 1.
 Pfam: PF01506; HCV_NS5a; 1.
 Pfam: PF00271; helicase_C; 1.
 Pfam: PF00998; Viral_RdRP; 1.
 DR PRODOM: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; GEDPoc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 MATRIX PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;
 Query Match 89.0%; Score 137; DB 1; Length 3011;
 Best Local Similarity 89.3%; Pred. No. 4.6e-12;
 Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 PKPQKTKRNTYRRPQDVKFPGGQIVG 28
 Db 5 PKPQKKNKNTNRRPQDVKFPGGQIVG 32
 RESULT 14
 FURI_HUMAN
 ID FURI_HUMAN STANDARD; PRT; 794 AA.
 AC P09958; Q14336;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Furin precursor (EC 3.4.21.75) (Paired basic amino acid residue
 DE cleaving enzyme) (PACE) (Dibasic processing enzyme).
 GN FURIN OR PACE OR FUR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=90175002; PubMed=2408021;
 RA van den Ouweland A.M.W., van Duijnhoven H.L.P., Keizer G.D.,
 RA Dorssers L.C.J., van de Ven W.J.M.;
 RA "Structural homology between the human fur gene product and the
 RT subtilisin-like protease encoded by yeast KEX2.";
 RL Nucleic Acids Res. 18:664-664(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9131735; PubMed=1713771;
 RA Barr P.J., Mason O.B., Landsberg K.E., Wong P.A., Kiefer M.C.,
 RA Brake A.J.;
 RA "cDNA and gene structure for a human subtilisin-like protease with
 RT cleavage specificity for paired basic amino acid residues.";
 RL DNA Cell Biol. 10:319-328(1991).
 RN [3]
 RP SEQUENCE OF 296-794 FROM N.A.
 RX MEDLINE=87053858; PubMed=3023061;
 RA Roebroek A.J.M., Schaiken J.A., Leunissen J.A.M., Onnekink C.,
 RA Bloemers H.P.J., van de Ven W.J.M.;
 RA "Evolutionary conserved close linkage of the c-fes/fps proto-oncogene
 RT and genetic sequences encoding a receptor-like protein.";
 RL EMBO J. 5:2197-2202(1986).
 RN [4]
 RP 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.
 RX MEDLINE=94291619; PubMed=8020465;
 RA Slezep R.J., Creemers J.W.M., van de Ven W.J.M.;
 RA "Homology modelling of the catalytic domain of human furin. A model
 RT for the eukaryotic subtilisin-like proprotein convertases.";
 RL Eur. J. Biochem. 222:255-266(1994).
 RN [5]
 RP PROCESSING.
 RX MEDLINE=92332543; PubMed=1629222;
 RA Leduc R., Molloy S.S., Thorne B.A., Thomas G.;
 RA "Activation of human furin precursor processing
 RT an intramolecular autoproteolytic cleavage.";
 RL J. Biol. Chem. 267:14304-14308(1992).
 CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTIDASE
 CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
 CC CATALYZING AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-1-zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
 CC complement component C3 and von Willebrand factor from their
 CC respective precursors.
 CC -1- COFACTOR: CALCIUM-DEPENDENT.

CC -1- ENZYME REGULATION: COULD BE INHIBITED BY THE NOT SECONDLY CLEAVED
 CC PROPEPTIDE.
 CC -1- SUBCELLULAR LOCATION: SEEMS TO BE LOCALIZED INTRACELLULARLY TO THE
 CC TRANS GOLGI NETWORK. PROPEPTIDE CLEAVAGE IS A PREREQUISITE FOR
 CC EXIT OF FURIN MOLECULES OUT OF THE ENDOPLASMIC RETICULUM (ER).
 CC SECOND CLEAVAGE IN THE PROPEPTIDE OCCUR IN THE TRANS GOLGI NETWORK
 CC (TGN), WHICH IS FOLLOWED BY THE RELEASE OF THE PROPEPTIDE BOUND TO
 CC FURIN AND THE ACTIVATION OF FURIN.
 CC -1- TISSUE SPECIFICITY: SEEMS TO BE EXPRESSED UBQUITOUSLY.
 CC -1- DOMAIN: CONTAINS A CYTOPLASMIC DOMAIN RESPONSIBLE FOR ITS TGN
 CC LOCALIZATION AND RECYCLING FROM THE CELL SURFACE.
 CC -1- PTM: THE PROPEPTIDE IS AUTOCATALYTICALLY REMOVED THROUGH AN
 CC INTRAMOLECULAR CLEAVAGE PROBABLY IN THE ENDOPLASMIC RETICULUM
 CC (ER). IN THE TGN THE SECOND CLEAVAGE IN THE PROPEPTIDE COULD LEAD
 CC TO THE ACTIVATION OF FURIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 homo B/P domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X17094; CAA34948.1; -;
 CC DR EMBL: X04329; CAA27860.1; -;
 CC DR EMBL: A06939; CAA00605.1; -;
 CC DR PIR: A39552; KXHUF.
 CC DR HSP: Q99405; IMPT.
 CC DR MEROPS: S08-071; -;
 CC DR Genev: HGNC:8568; FURIN.
 CC DR MIM: 136950; -;
 CC DR GO: GO:0005794; C:Golgi apparatus; TAS.
 CC DR GO: GO:0004276; F:furin activity; TAS.
 CC DR GO: GO:0007267; P:cell-cell signaling; TAS.
 CC DR GO: GO:0006508; P:proteolysis and peptidolysis; TAS.
 CC DR InterPro: IPR006212; Furin_repeat.
 CC DR InterPro: IPR002884; P_domain.
 CC DR Pfam: PF01483; P_protein; PARTIAL.
 CC DR Pfam: PF00082; Peptidase_S8; 1.
 CC DR PRINTS: PR00723; SUBTILISIN.
 CC DR ProDom: PD000717; P_domain; 1.
 CC DR SMART: SM00261; FU; 2.
 CC DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 CC DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 CC DR PROSITE: PS00138; SUBTILASE_SER; 1.
 CC DR Hydrolase; Serine protease; Transmembrane; Glycoprotein; Signal;
 CC -----
 CC SIGNAL 1 24 POTENTIAL.
 CC PROPEP 25 107
 CC CHAIN 108 794 FURIN.
 CC DOMAIN 556 705 CYS-RICH.
 CC TRANSMEM 716 738 POTENTIAL.
 CC ACT_SITE 153 153 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 194 194 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC DISULFID 211 360 POTENTIAL.
 CC DISULFID 303 333 POTENTIAL.
 CC CARBOHYD 387 387 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 440 440 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 553 553 N-LINKED (GLCNAC...) (POTENTIAL).
 CC SITE 70 75 CLEAVAGE (SECOND AUTO-).
 CC SITE 104 107 CLEAVAGE (FIRST AUTO-).
 CC SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).
 CC SITE 759 762 CELL SURFACE SIGNAL.
 CC SITE 773 779 TRANS GOLGI NETWORK SIGNAL.
 CC SEQUENCE 794 AA: 86678 MW: 10C44DD5892EF85D CRC64;
 CC -----
 CC Query Match 39.08; Score 60; DB 1; Length 794;
 CC Best Local Similarity 55.6%; Pred. No. 0.41;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Qy 4 QRKTKNTYRRPODVKFP 21
 Db 102 KRTRKRDVYQETDPKFP 119
 :|||: | : | |||
 102 KRTRKRDVYQETDPKFP 119
 RESULT 15
 FURIN_MOUSE
 ID FURIN_MOUSE STANDARD; PRT; 793 AA.
 AC P23188;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Furin precursor (EC 3.4.21.75) (paired basic amino acid residue
 DE cleaving enzyme) (PACE) (Dibasic processing enzyme).
 GN FURIN OR FUR OR PCSK3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91093035; PubMed=2266110;
 RA Hattuzawa K., Hosaka M., Nakagawa T., Nagase M., Shoda A.,
 RA Murakami K., Nakayama K.;
 RT "Structure and expression of mouse furin, a yeast Kex2-related
 RT protease. Lack of processing of coexpressed prorenin in GH4C1
 cells.";
 RL J. Biol. Chem. 265:22075-22078(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Creemers J.W.M., Roebroek A.J.M., van den Ouweland A.M.W.,
 RA van Duifjnhoven H.L.P., van de Ven W.J.M.;
 RT "Cloning and functional expression of a 4.3 kbp mouse fur cDNA:
 RT evidence for differential expression.";
 RL Life Sci. Adv. (Mol. Biol.) 11:127-138(1992).
 CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBQUITOUS ENDOPEPTASE
 CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
 CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their
 CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa
 CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
 CC complement component C3 and von Willebrand factor from their
 CC respective precursors.
 CC -1- COFACTOR: CALCIUM-DEPENDENT (BY SIMILARITY).
 CC -1- ENZYME REGULATION: COULD BE INHIBITED BY THE NOT SECONDLY CLEAVED
 CC PROPEPTIDE.
 CC -1- SUBCELLULAR LOCATION: SEEMS TO BE LOCALIZED INTRACELLULARLY TO THE
 CC TRANS GOLGI NETWORK. PROPEPTIDE CLEAVAGE IS A PREREQUISITE FOR
 CC EXIT OF FURIN MOLECULES OUT OF THE ENDOPLASMIC RETICULUM (ER).
 CC SECOND CLEAVAGE IN THE PROPEPTIDE OCCUR IN THE TRANS GOLGI NETWORK
 CC (TGN), WHICH IS FOLLOWED BY THE RELEASE OF THE PROPEPTIDE BOUND TO
 CC FURIN AND THE ACTIVATION OF FURIN.
 CC -1- TISSUE SPECIFICITY: SEEMS TO BE EXPRESSED UBQUITOUSLY.
 CC -1- DOMAIN: CONTAINS A CYTOPLASMIC DOMAIN RESPONSIBLE FOR ITS TGN
 CC LOCALIZATION AND RECYCLING FROM THE CELL SURFACE.
 CC -1- PTM: THE PROPEPTIDE IS AUTOCATALYTICALLY REMOVED THROUGH AN
 CC INTRAMOLECULAR CLEAVAGE PROBABLY IN THE ENDOPLASMIC RETICULUM
 CC (ER). IN THE TGN THE SECOND CLEAVAGE IN THE PROPEPTIDE COULD LEAD
 CC TO THE ACTIVATION OF FURIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 homo B/P domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X54056; CAA37988.1; -
 DR EMBL; L26489; AAA37643.1; -
 DR PIR; A23679; KXMF.
 DR HSSP; Q99405; IMPT.
 DR MEROPS; S08.071; -
 DR MGD; MGI:97513; Furin.
 DR InterPro; IPR006212; Furin.repeat.
 DR InterPro; IPR002884; P_domain.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF01483; P_protein; PARTIAL.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PRO0723; SUBTILISIN.
 DR ProDom; PD00717; P_domain; 1.
 DR SMART; SM00261; FU; 2.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Glycoprotein; Signal;
 KW zymogen; Calcium.
 FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 107 BY SIMILARITY.
 FT CHAIN 108 793 FURIN.
 FT DOMAIN 536 705 CYS-RICH.
 FT TRANSMEM 715 735 POTENTIAL.
 FT ACT_SITE 153 153 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 194 194 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 211 360 POTENTIAL.
 FT DISULFID 303 333 POTENTIAL.
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 70 75 CLEAVAGE (SECOND AUTO-).
 FT SITE 104 107 CLEAVAGE (FIRST AUTO-).
 FT SITE 758 761 CELL SURFACE SIGNAL.
 FT SITE 772 778 TRANS GOLGI NETWORK SIGNAL.
 FT SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).
 FT CONFLICT 746 746 M -> V (IN REF. 2).
 SQ SEQUENCE 793 AA; 86804 MW; 5F121C3DE2E1A42D CRC64;

Query Match 35.7%; Score 55; DB 1; Length 793;
 Best Local Similarity 50.0%; Pred. No. 2.3;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 4 QKTKRNTYRRPDVKFP 21
 :|: ||: | : ||||
 DB 102 KRAKRDVYQETDPKFP 119

.rch completed: August 7, 2003, 11:19:58
 .o time : 5.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds

(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-25
Perfect score: 154
Sequence: 1 PRPQRTKNTYRRPDVFKPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL23:*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	105	12	P90364
2	154	100.0	106	12	O81831
3	154	100.0	125	12	O9PXN3
4	154	100.0	131	12	O68575
5	154	100.0	3010	12	O91AU0
6	154	100.0	3010	12	O81989
7	154	100.0	3010	12	O81541
8	150	97.4	109	12	O81807
9	146	94.8	45	12	O68307
10	146	94.8	45	12	O68308
11	146	94.8	45	12	O68310
12	146	94.8	60	12	O8JYS2
13	146	94.8	61	12	O8JYS3
14	146	94.8	73	12	O8JYR4
15	146	94.8	100	12	O8QP85
16	146	94.8	100	12	O8QP87

17	146	94.8	100	12	O8QP72	O8qp72 hepatitis c
18	146	94.8	100	12	O8QP71	O8qp71 hepatitis c
19	146	94.8	100	12	O8QP74	O8qp74 hepatitis c
20	146	94.8	100	12	O8QP80	O8qp80 hepatitis c
21	146	94.8	100	12	O8QP77	O8qp77 hepatitis c
22	146	94.8	100	12	O8QP84	O8qp84 hepatitis c
23	146	94.8	100	12	O8QP78	O8qp78 hepatitis c
24	146	94.8	100	12	O8QP75	O8qp75 hepatitis c
25	146	94.8	100	12	O8QP79	O8qp79 hepatitis c
26	146	94.8	100	12	O8QP81	O8qp81 hepatitis c
27	146	94.8	100	12	O8QP83	O8qp83 hepatitis c
28	146	94.8	100	12	O8QP76	O8qp76 hepatitis c
29	146	94.8	100	12	O8QP86	O8qp86 hepatitis c
30	146	94.8	109	12	O81340	O81340 hepatitis c
31	146	94.8	114	12	O8QMJ4	O8qmj4 hepatitis c
32	146	94.8	114	12	O68892	O68892 hepatitis c
33	146	94.8	114	12	O68893	O68893 hepatitis c
34	146	94.8	119	12	O8CX2	O8bcx2 hepatitis c
35	146	94.8	119	12	O8CX1	O8bcx1 hepatitis c
36	146	94.8	119	12	O8BCX0	O8bcx0 hepatitis c
37	146	94.8	119	12	O8BCW9	O8bcw9 hepatitis c
38	146	94.8	119	12	O8BCW8	O8bcw8 hepatitis c
39	146	94.8	119	12	O8BCW7	O8bcw7 hepatitis c
40	146	94.8	119	12	O8BCW6	O8bcw6 hepatitis c
41	146	94.8	119	12	O8BCW4	O8bcw4 hepatitis c
42	146	94.8	122	12	O8QRJ1	O8qrj1 hepatitis c
43	146	94.8	122	12	O8QMI6	O8qm16 hepatitis c
44	146	94.8	122	12	O8QMI8	O8qm18 hepatitis c
45	146	94.8	122	12	O8QRJ4	O8qrj4 hepatitis c

ALIGNMENTS

RESULT 1

P90364 ID P90364 PRELIMINARY; PRT; 105 AA.
AC P90364;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohno T., Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR
method.";
RL Thesis (1995), Nagoya City University Medical School.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305314; PubMed=8712927;
RA Ohno T., Mizokami M., Saleh M.G., Orito E., Ohba K.-I., Wu R.-R.,
Ra Koide T., Tibbs C.J., Nouri-Aria K.T., Tokudome S., Williams R.;
RT "Usefulness and limitation of phylogenetic analysis for hepatitis C
virus core region: application to isolates from Egyptian and Yemeni
patients.";
RL Arch. Virol. 141:1101-1113(1996).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; D82034; E8A11519.1; -;
DR Interpro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
FT NON_TER 1 105
FT NON_TER 105 105
SQ SEQUENCE 105 AA; 11870 MW; C7BA40B284025A49 CRC64;

Query Match 100.0%; Score 154; DB 12; Length 105;
 Best Local Similarity 100.0%; Pred. No. 9.8e-16;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
 |||||
 DB 4 PKPQKTKRNTYRRPDQVKFPGGGQIVG 31

RESULT 2

Q81831 PRELIMINARY; PRT; 106 AA.

ID Q81831
 AC Q81831
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE (isolate USA8) genomic RNA (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;

[1]
 SEQUENCE FROM N.A.

RC STRAIN-US8;
 RA Ohno T., Mizokami M.;
 RT "Determination of nine genotypes of hepatitis C virus using PCR method."

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RT -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL; D49465; BAA08439.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR Pfam; PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER
 KX 106
 SQ SEQUENCE 106 AA; 12001 MW; 25D0D5414B3EA9DC CRC64;

Query Match 100.0%; Score 154; DB 12; Length 106;
 Best Local Similarity 100.0%; Pred. No. 9.9e-16;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
 |||||
 DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 3

Q9PXN3

PRELIMINARY; PRT; 125 AA.

AC Q9PXN3
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE E2/NS1 protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus type 2.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=40271;

[1]
 SEQUENCE FROM N.A.

RC MEDLINE-94351179; PubMed=7520922;
 RA Nakazawa T., Kato N., Ohkoshi S., Shibuya A., Shimotohno K.;
 RL J. Hepatol. 20:623-629(1994).

RT -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).

DR InterPro; IPR002522; HCV_capsid.
 DR Pfam; PF01543; HCV_capsid; 1.
 KW Polyprotein.
 FT NON_TER
 KX 125 AA; 14220 MW; D945CB60661797E3 CRC64;

Query Match 100.0%; Score 154; DB 12; Length 125;
 Best Local Similarity 100.0%; Pred. No. 1.2e-15;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
 |||||
 DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 4

Q68575 PRELIMINARY; PRT; 191 AA.

ID Q68575
 AC Q68575
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Core protein (Genome polyprotein) (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;

[1]
 SEQUENCE FROM N.A.

RC STRAIN-BB51;
 RA Songsivilai S., Dharakul T., Kunkitti R., Thepthal C.;
 RT "Molecular cloning and expression of hepatitis C virus core protein
 and production of monoclonal antibodies to the recombinant protein.";
 RL Asian Pac. J. Allergy Immunol. 14:0-0(1996).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS;
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).

DR EMBL; U55284; AAB00215.1; -;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV core.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 KW Polyprotein.
 FT NON_TER
 KX 191
 SQ SEQUENCE 191 AA; 20840 MW; 4AAE63444D8329E2 CRC64;

Query Match 100.0%; Score 154; DB 12; Length 191;
 Best Local Similarity 100.0%; Pred. No. 1.9e-15;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
 |||||
 DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32

RESULT 5

Q91AU0 PRELIMINARY; PRT; 3010 AA.

ID Q91AU0
 AC Q91AU0
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.

OX NCBI_TaxID=11103;

[1]
 SEQUENCE FROM N.A.

RC STRAIN-HCV-S1;
 RX MEDLINE-21440119; PubMed=11556407;

RT Lim S.P., Khu Y.L., Hong W.J., Tay A., Ting A.E., Lim S.G., Tan Y.H.;
 "Identification and molecular characterization of the complete genome
 of a Singapore isolate of hepatitis C virus: sequence comparison with
 other strains and phylogenetic analysis.";
 RL Virus Genes 23:89-95(2001).
 RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN-HCV-SI;
RL Lim S.P.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AF356827; AAL00300.1; -.
DR InterPro: IPR000345; Cyt_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002531; HCV_NS2.
DR InterPro: IPR002518; HCV_NS3.
DR InterPro: IPR004109; HCV_NS4a.
DR InterPro: IPR000745; HCV_NS4b.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326793 MW; 3DB9304314F9F795 CRC64;

Query Match 100.0%; Score 154; DB 12; Length 3010;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQRTKRTYRRPDVQKFPGGQIVG 28
|||||
5 PKPQRTKRTYRRPDVQKFPGGQIVG 32

RESULT 6
ID Q81989 PRELIMINARY; PRT; 3010 AA.
AC Q81989;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E1 and E2/NSI envelope glycoprotein (Genome polyprotein).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-N;
RA Zheng W.-Y.Z.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-N;
RA Zheng W.Z.;
RT "Genotype identification of hepatitis c virus (HCV) isolated from a

```

```

RT single Japanese carrier in Nagasaki prefecture and genome analysis of
RT E1 and E2/NSI envelope glycoprotein regions.";
RL Jpn. J. Trop. Med. Hyg. 22:169-177(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D63857; BAA09919.1; -.
DR HSP: P26663; LJXP.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002531; HCV_NS2.
DR InterPro: IPR002518; HCV_NS3.
DR InterPro: IPR004109; HCV_NS4a.
DR InterPro: IPR000745; HCV_NS4b.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW APP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327503 MW; C7BDB38169D6E3CF CRC64;

Query Match 100.0%; Score 154; DB 12; Length 3010;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPQRTKRTYRRPDVQKFPGGQIVG 28
|||||
5 PKPQRTKRTYRRPDVQKFPGGQIVG 32

RESULT 7
ID Q81541 PRELIMINARY; PRT; 3010 AA.
AC Q81541;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
DE Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JT';
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T.,
RA Hijikata M., Shimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";

```

```

Virus Res. 23:39-53(1992).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; D11355; BAA18894.1; -
DR HSP: P26663; LJPX.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_NS5b.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
Pfam; PF01543; HCV_capsid; 1.
Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_env; 1.
Pfam; PF01560; HCV_NS1; 1.
Pfam; PF01538; HCV_NS2; 1.
Pfam; PF02907; HCV_NS3; 1.
Pfam; PF01006; HCV_NS4a; 1.
Pfam; PF01001; HCV_NS4b; 1.
Pfam; PF01506; HCV_NS5a; 1.
Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RORP_POSITIVE; 1.
DR PROSITE; PS50521; RORP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN PUT_NS5.
SQ SEQUENCE 3010 AA; 326564 MW; 05F0B2102CF9DD9D CRC64;

Query Match 100.0%; Score 154; DB 12; Length 3010;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 PKPQKTKRNTYRRPQDVKFGGQIVG 28
|||||
5 PKPQKTKRNTYRRPQDVKFGGQIVG 32

RESULT 8
O81807 PRELIMINARY; PRT; 109 AA.
ID O81807;
AC O81807;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE (isolate SL1) genomic RNA (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SLI;
RA Ohno T.; Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR
methdod.";

Virus Res. 23:39-53(1992).
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; D11355; BAA18894.1; -
DR HSP: P26663; LJPX.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_NS5b.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
Pfam; PF01543; HCV_capsid; 1.
Pfam; PF01542; HCV_core; 1.
Pfam; PF01539; HCV_env; 1.
Pfam; PF01560; HCV_NS1; 1.
Pfam; PF01538; HCV_NS2; 1.
Pfam; PF02907; HCV_NS3; 1.
Pfam; PF01006; HCV_NS4a; 1.
Pfam; PF01001; HCV_NS4b; 1.
Pfam; PF01506; HCV_NS5a; 1.
Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RORP_POSITIVE; 1.
DR PROSITE; PS50521; RORP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN PUT_NS5.
SQ SEQUENCE 3010 AA; 326564 MW; 05F0B2102CF9DD9D CRC64;

Query Match 100.0%; Score 154; DB 12; Length 3010;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 PKPQKTKRNTYRRPQDVKFGGQIVG 28
|||||
5 PKPQKTKRNTYRRPQDVKFGGQIVG 32

RESULT 8
O81807 PRELIMINARY; PRT; 109 AA.
ID O81807;
AC O81807;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE (isolate SL1) genomic RNA (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SLI;
RA Ohno T.; Mizokami M.;
RT "Determination of nine genotypes of hepatitis C virus using PCR
methdod.";

```

```

Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; D49459; BAA21059.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 109
SQ SEQUENCE 109 AA; 12318 MW; 9999DBD249BF93H0 CRC64;

Query Match 97.4%; Score 150; DB 12; Length 109;
Best Local Similarity 96.4%; Pred. No. 4.1e-15;
Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPQDVKFGGQIVG 28
|||||
5 PKPQKTKRNTYRRPQDVKFGGQIVG 32

DB 5 PKPQKTKRNTYRRPQDVKFGGQIVG 32

RESULT 9
Q68307 PRELIMINARY; PRT; 45 AA.
ID Q68307;
AC Q68307;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB21;
RA Songvilai S.; Kanistanon D.; Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23745; AAA65052.1; -
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 45
SQ SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 45;
Best Local Similarity 96.4%; Pred. No. 6.5e-15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPQDVKFGGQIVG 28
|||||
5 PKPQKTKRNTYRRPQDVKFGGQIVG 32

DB 5 PKPQKTKRNTYRRPQDVKFGGQIVG 32

RESULT 10
Q68308 PRELIMINARY; PRT; 45 AA.
ID Q68308;
AC Q68308;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]

```



```
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-BB38;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; U23746; AAA65053.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein. 45 45
FT NON_TER
SQ SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 45;
Best Local Similarity 96.4%; Pred. No. 6.5e-15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 PKPQRTKRTYRRPDVKFPGGGQIVG 28
5 PKPQRTKRTYRRPDVKFPGGGQIVG 32
|||||
|||||

RESULT 11
Q68310 PRELIMINARY; PRT; 45 AA.
AC Q68310;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HCV-C94009;
RA Songsivilai S., Kanistanon D., Kunkitti R.;
RT "Identification and characterisation of Thai isolates of hepatitis C
virus.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; U23748; AAA65055.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein. 45 45
FT NON_TER
SQ SEQUENCE 45 AA; 5014 MW; CC527167096DDAF6 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 45;
Best Local Similarity 96.4%; Pred. No. 6.5e-15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTYRRPDVKFPGGGQIVG 28
5 PKPQRTKRTYRRPDVKFPGGGQIVG 32
|||||
|||||

RESULT 12
Q8JYS2 PRELIMINARY; PRT; 60 AA.
AC Q8JYS2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
```

```
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG253;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF506613; AAM33389.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein. 60 60
FT NON_TER
SQ SEQUENCE 60 AA; 6725 MW; ACCE7D9C8B90299A CRC64;

Query Match 94.8%; Score 146; DB 12; Length 60;
Best Local Similarity 96.4%; Pred. No. 8.9e-15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTYRRPDVKFPGGGQIVG 28
5 PKPQRTKRTYRRPDVKFPGGGQIVG 32
|||||
|||||

RESULT 13
Q8JYS3 PRELIMINARY; PRT; 61 AA.
AC Q8JYS3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG256;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
EMBL; AF506612; AAM33388.1; -
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
KW Polyprotein. 1 >61 CORE PROTEIN.
FT CHAIN 61 61
FT NON_TER 61 61
SQ SEQUENCE 61 AA; 6881 MW; B92CCE7D9C8B9029 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 61;
Best Local Similarity 96.4%; Pred. No. 9e-15;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRTYRRPDVKFPGGGQIVG 28
5 PKPQRTKRTYRRPDVKFPGGGQIVG 32
|||||
|||||

RESULT 14
Q8JYR4 PRELIMINARY; PRT; 73 AA.
AC Q8JYR4;
```

Search completed: August 7, 2003, 11:18:58
Job time : 26.6364 secs

```
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RIG291;
RA Shustov A.V., Gavrilova I.V., Netesov S.V.;
RT "Genetic variability of hepatitis C virus in Western Siberia.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF506621; AA33397.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam: PF01543; HCV_capsid; 1.
DR Polyprotein.
RA NON_TER 1
FT NON_TER 73 73
SQ SEQUENCE 73 AA; 8201 MW; 6EDC082DAE8CB796 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 73;
Best Local Similarity 96.4%; Pred. No. 1.1e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
DB 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28

RESULT 15
Q8QP85 PRELIMINARY; PRT; 100 AA.
AC Q8QP85;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-602;
RA MEDLINE-21904745; PubMed-11907242;
RT Kalinina O., Nordier H., Mukomolov S., Magnus L.O.;
RT "A natural intergenotypic recombinant of hepatitis C virus identified
RT in St. Petersburg.";
RL J. Virol. 76:4034-4043(2002).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AY070180; AAL58594.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
FT NON_TER 100 100
SQ SEQUENCE 100 AA; 11236 MW; 12AD90F5F5885AE6 CRC64;

Query Match 94.8%; Score 146; DB 12; Length 100;
Best Local Similarity 96.4%; Pred. No. 1.5e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDQVKFPGGGQIVG 28
DB 5 PKPQKTKRNTYRRPDQVKFPGGGQIVG 32
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(Without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-25

Perfect score: 154

Sequence: 1 PKPQKTRNTYRRPDVKFPGGGOIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues
total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/8A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/8B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	28	3	US-08-921-887-25
2	154	100.0	450	4	US-08-635-886C-191
3	154	100.0	450	4	US-08-635-886C-192
4	146	94.8	28	3	US-08-921-887-23
5	146	94.8	191	2	US-08-290-665A-187
6	146	94.8	191	2	US-08-290-665A-188
7	146	94.8	191	2	US-08-290-665A-190
8	146	94.8	191	5	PCT-US95-10398-187
9	146	94.8	191	5	PCT-US95-10398-188
10	146	94.8	191	5	PCT-US95-10398-190
11	146	94.8	319	3	US-08-836-075A-44
12	146	94.8	319	4	US-08-635-886C-230
13	145	94.2	34	3	US-08-380-160-6
14	145	94.2	43	4	US-09-020-846-36
15	145	94.2	44	3	US-08-380-160-2
16	145	94.2	44	4	US-09-389-756-1
17	145	94.2	45	3	US-08-380-160-1
18	145	94.2	61	1	US-07-946-054-9
19	145	94.2	61	1	US-08-083-947-23
20	145	94.2	61	1	US-08-530-550-3
21	145	94.2	61	1	US-08-262-037-26
22	145	94.2	61	5	PCT-US93-08638-9
23	145	94.2	61	5	PCT-US94-07088-23
24	145	94.2	61	5	PCT-US95-13660-3
25	145	94.2	74	3	US-08-836-075A-10
26	145	94.2	74	4	US-08-635-886C-198
27	145	94.2	100	4	US-08-635-886C-232

28 145 94.2 108 3 US-08-836-075A-14
29 145 94.2 115 1 US-08-324-977-8
30 145 94.2 115 2 US-08-384-616-8
31 145 94.2 115 2 US-08-904-686A-8
32 145 94.2 115 3 US-09-315-850-8
33 145 94.2 123 2 US-08-501-195-2
34 145 94.2 137 3 US-08-836-075A-46
35 145 94.2 138 3 US-08-836-075A-60
36 145 94.2 140 2 US-08-501-195-1
37 145 94.2 154 3 PCT-US95-13552-2
38 145 94.2 154 5 PCT-US95-13552-2
39 145 94.2 190 1 US-07-681-701-16
40 145 94.2 190 3 US-08-078-271B-1
41 145 94.2 191 2 US-08-290-665A-155
42 145 94.2 191 2 US-08-290-665A-156
43 145 94.2 191 2 US-08-290-665A-157
44 145 94.2 191 2 US-08-290-665A-158
45 145 94.2 191 2 US-08-290-665A-159

ALIGNMENTS

RESULT 1
US-08-921-887-25
; Sequence 25, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; METHOD FOR MAKING THE SA
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-25

Query Match 100.0%; Score 154; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.9e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPQKTKRNTYRRPQDVKFPGGGQIVG 28
Db 1 PKPQKTKRNTYRRPQDVKFPGGGQIVG 28

RESULT 2

US-08-635-886C-191
; Sequence 191, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-191

Query Match 100.0%; Score 154; DB 4; Length 450;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPQKTKRNTYRRPQDVKFPGGGQIVG 28
Db 5 PKPQKTKRNTYRRPQDVKFPGGGQIVG 32

RESULT 3

US-08-635-886C-192
; Sequence 192, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635.886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 192
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-192

Query Match 100.0%; Score 154; DB 4; Length 450;
Best Local Similarity 100.0%; Pred. No. 3.3e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKPQKTKRNTYRRPQDVKFPGGGQIVG 28
Db 5 PKPQKTKRNTYRRPQDVKFPGGGQIVG 32

RESULT 4

US-08-921-887-23
; Sequence 23, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUYAKOV, YURI E.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAM
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
US-08-921-887-23

Query Match 94.8%; Score 146; DB 3; Length 28;
Best Local Similarity 96.4%; Pred. No. 2.7e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKPQKTKRNTYRRPQDVKFPGGGQIVG 28
Db 1 PKPQKTKRNTYRRPQDVKFPGGGQIVG 28

RESULT 5

US-08-290-665A-187
; Sequence 187, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURK, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK

STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK10
US-08-290-665A-187

Query Match 94.8%; Score 146; DB 2; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.9e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTYRRPQDVKFPGGQIVG 28
|||||
Db 5 PKPQKTKRNTYRRPQDVKFPGGQIVG 32

RESULT 6
US-08-290-665A-188
Sequence 188, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
US-08-290-665A-188

Query Match 94.8%; Score 146; DB 2; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.9e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKPQKTKRNTYRRPQDVKFPGGQIVG 28
|||||
Db 5 PKPQKTKRNTYRRPQDVKFPGGQIVG 32

RESULT 7
US-08-290-665A-190
Sequence 190, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 190:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
US-08-290-665A-190

Query Match 94.8%; Score 146; DB 2; Length 191;

Best Local Similarity 96.4%; Pred. No. 1.9e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKFFGGGOIVG 28
|||||
Db 5 PKPQKTKRNTYRRPQDVKFFGGGOIVG 32
|||||

RESULT 8

PCT-US95-10398-187

; Sequence 187, Application PC/TUS9510398

; GENERAL INFORMATION:

; APPLICANT: BUKH, J., MILLER, R.H. AND

; APPLICANT: PURCELL, R.H.

; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

; NUMBER OF SEQUENCES: 263

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10398

; FILING DATE: 15-AUG-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/086,428

; FILING DATE: 29 JUNE 1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/290/665

; FILING DATE: 15 AUGUST 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: RICHARD W. BORK

; REGISTRATION NUMBER: 36,459

; REFERENCE/DOCKET NUMBER: 2026-4116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 187:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 191 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; ORIGINAL SOURCE:

; ORGANISM: homosapiens

; INDIVIDUAL ISOLATE: HK10

; PCT-US95-10398-187

Query Match 94.8%; Score 146; DB 5; Length 191;

Best Local Similarity 96.4%; Pred. No. 1.9e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKFFGGGOIVG 28
|||||
Db 5 PKPQKTKRNTYRRPQDVKFFGGGOIVG 32
|||||

RESULT 9

PCT-US95-10398-188

; Sequence 188, Application PC/TUS9510398

; GENERAL INFORMATION:

; APPLICANT: BUKH, J., MILLER, R.H. AND

; APPLICANT: PURCELL, R.H.

; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

; NUMBER OF SEQUENCES: 263

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10398

; FILING DATE: 15-AUG-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/086,428

; FILING DATE: 29 JUNE 1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/290/665

; FILING DATE: 15 AUGUST 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: RICHARD W. BORK

; REGISTRATION NUMBER: 36,459

; REFERENCE/DOCKET NUMBER: 2026-4116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 188:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 191 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; ORIGINAL SOURCE:

; ORGANISM: homosapiens

; INDIVIDUAL ISOLATE: S52

; PCT-US95-10398-188

Query Match

Best Local Similarity 96.4%; Score 146; DB 5; Length 191;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKFFGGGOIVG 28
|||||
Db 5 PKPQKTKRNTYRRPQDVKFFGGGOIVG 32
|||||

RESULT 10

PCT-US95-10398-190

; Sequence 190, Application PC/TUS9510398

; GENERAL INFORMATION:

; APPLICANT: BUKH, J., MILLER, R.H. AND

; APPLICANT: PURCELL, R.H.

; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

; NUMBER OF SEQUENCES: 263

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

```

STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 190:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
PCT-US95-10398-190

Query Match          94.8%; Score 146; DB 5; Length 191;
Best Local Similarity 96.4%; Pred No. 1,9e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQRKTKRNTYRRPDVKFPGGGQIVG 28
Db      5 PKPQRKTKRNTIRRPDVKFPGGGQIVG 32

BULT 11
-08-836-075A-44
: Sequence 44, Application US/08836075A
: Patent No. 6180768
: GENERAL INFORMATION:
: APPLICANT: MAERTENS, GEERT
: APPLICANT: STUYVER, LIEVEN
: TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
: TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
: TITLE OF INVENTION: AGENTS
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE & DURKEE
: STREET: P.O. BOX 4433
: CITY: HOUSTON
: STATE: TEXAS
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Microsoft Word 6.0 / ASCII text output
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/836,075A

```

```

; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA: PCT/EP95/04155
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-075A-44

Query Match 94.8%; Score 146; DB 3; Length 319;
Best Local Similarity 96.4%; Pred.No. 3.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRTKRNTRYRRPDVKFPGGGQIVG 28
Db 5 PKPQRTKRNTRYRRPDVKFPGGGQIVG 32
|||||
|||||

RESULT 12
US-08-635-886C-230
; Sequence 230, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 230
; LENGTH: 319
; TYPE: PRT
; ORGANISM: hepatitis C virus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (144)..(144)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (156)..(157)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (161)..(161)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (167)..(167)

```

SEQUENCE CHARACTERISTICS:

Query Match	Best Local Similarity	Best Local Similarity
94.2%	96.4%	96.4%

Search completed: August 7, 2003, 11:23:50
Job time : 10.6364 secs

QY 1 PKPQKTKRNTYRRPDVKFPGGGQIVG 28
|||||
Db 5 PKPQKTKRNTNRRPDVKFPGGGQIVG 32
|||||

RESULT 15
US-08-380-160-2
; Sequence 2, Application US/08380160
; Patent No. 6235284
; GENERAL INFORMATION:
; APPLICANT: DALBON, Pascal
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
; TITLE OF INVENTION: FOR DETECTING THE LATTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,160
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,471
; FILING DATE: 06-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPH 28682
; TELEPHONE: (703)836-6400
; TELEFAX: (703)836-2787
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Human Hepatitis C Virus
; STRAIN: H77
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..44
; OTHER INFORMATION: /note= "N-terminal sequence of the
; OTHER INFORMATION: protein of the nucleocapside or CORE protein of
; OTHER INFORMATION: the human hepatitis C virus"
US-08-380-160-2

Query Match 94.2%; Score 145; DB 3; Length 44;
Best Local Similarity 96.4%; Pred. No. 5,9e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKTKRNTYRRPDVKFPGGGQIVG 28
|||||
Db 4 PKPQKTKRNTNRRPDVKFPGGGQIVG 31
|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146a-25
Perfect score: 154
Sequence: 1 PKPQRTKNTYRPPQDVKPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues
Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	94.8	319	10	US-09-851-138-44
2	145	94.2	44	15	US-10-367-677-1
3	145	94.2	74	10	US-09-851-138-10
4	145	94.2	91	9	US-09-758-308-1
5	145	94.2	97	10	US-09-756-875-8
6	145	94.2	103	10	US-09-921-397-77
7	145	94.2	108	10	US-09-851-138-14
8	145	94.2	113	10	US-09-921-397-78
9	145	94.2	137	10	US-09-851-138-46
10	145	94.2	138	10	US-09-851-138-60
11	145	94.2	166	11	US-09-899-046-152
12	145	94.2	166	11	US-09-878-281-152
13	145	94.2	169	11	US-09-899-046-42
14	145	94.2	169	11	US-09-899-046-44
15	145	94.2	169	11	US-09-878-281-42

16	145	94.2	169	11	US-09-878-281-44
17	145	94.2	182	10	US-09-929-955-2
18	145	94.2	182	14	US-10-104-966-2
19	145	94.2	191	11	US-09-194-949-3
20	145	94.2	318	10	US-09-851-138-76
21	145	94.2	319	10	US-09-851-138-12
22	145	94.2	319	10	US-09-851-138-18
23	145	94.2	319	11	US-09-899-046-50
24	145	94.2	319	11	US-09-899-046-52
25	145	94.2	319	11	US-09-899-046-54
26	145	94.2	319	11	US-09-899-046-144
27	145	94.2	319	11	US-09-878-281-50
28	145	94.2	319	11	US-09-878-281-52
29	145	94.2	319	11	US-09-878-281-54
30	145	94.2	319	11	US-09-878-281-144
31	145	94.2	809	10	US-09-973-025-50
32	145	94.2	809	11	US-09-899-303-50
33	145	94.2	809	11	US-09-995-808-50
34	145	94.2	809	11	US-09-995-860-50
35	145	94.2	2894	10	US-09-941-611-23
36	145	94.2	2894	15	US-10-044-995-23
37	145	94.2	2985	15	US-10-259-275-40
38	145	94.2	3011	9	US-09-742-659-4
39	145	94.2	3011	10	US-09-952-572-9
40	145	94.2	3011	10	US-09-929-955-1
41	145	94.2	3011	10	US-09-747-419-20
42	145	94.2	3011	11	US-09-891-894-3
43	145	94.2	3011	14	US-10-104-966-1
44	145	94.2	3011	15	US-10-259-275-20
45	145	94.2	3012	10	US-09-238-076-2

ALIGNMENTS

RESULT 1

US-09-851-138-44
; Sequence 44, Application US/09851138
; Publication NO. US20020183508A1
; GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN

AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433

CITY: HOUSTON

STATE: TEXAS

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,138

FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>

APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994

APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:
LENGTH: 319 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-851-138-44

Query Match 94.2%; Score 146; DB 10; Length 319;
Best Local Similarity 96.4%; Pred. No. 6.2e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 2
US-10-367-677-1
Sequence 1, Application US/10367677
Publication No. US20030118604A1
GENERAL INFORMATION:
APPLICANT: JOLIVET, MICHEL
APPLICANT: PENIN, FRANCOIS
APPLICANT: DALBON, PASCAL
APPLICANT: LADAVIERE, LAURENT
APPLICANT: LACOUX, XAVIER
TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
TITLE OF INVENTION: COMPOUNDS, AND USES FOR DETECTING, PREVENTING AND
TITLE OF INVENTION: TREATING AN HCV INFECTION
FILE REFERENCE: 103959
CURRENT APPLICATION NUMBER: US/10/367,677
CURRENT FILING DATE: 2003-02-19
PRIOR APPLICATION NUMBER: US/09/389,756
PRIOR FILING DATE: 1999-09-07
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/FR98/00442
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 1
LENGTH: 44
TYPE: PRT
ORGANISM: Hepatitis C virus
PUBLICATION INFORMATION:
AUTHORS: Ogata, N. et al.
TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 3392-3396
DATE: 1991
RELEVANT RESIDUES: 2 TO 45
US-10-367-677-1

Query Match 94.2%; Score 145; DB 15; Length 44;
Best Local Similarity 96.4%; Pred. No. 9.9e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
Db 4 PKPQKTKRNTYRRPQDVKPPGGQIVG 31

RESULT 3
US-09-851-138-10
Sequence 10, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS

NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-851-138-10

Query Match 94.2%; Score 145; DB 10; Length 74;
Best Local Similarity 96.4%; Pred. No. 1.7e-11;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 4
US-09-758-308-1
Sequence 1, Application US/09758308
Patent No. US20020090607A1
GENERAL INFORMATION:
APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C
TITLE OF INVENTION: PROTEINS
FILE REFERENCE: 14114.0349U2
CURRENT APPLICATION NUMBER: US/09/758,308
CURRENT FILING DATE: 2001-01-10
PRIOR APPLICATION NUMBER: 60/092,339
PRIOR FILING DATE: 1999-07-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 91
TYPE: PRT
ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 94.2%; Score 145; DB 9; Length 91;
Best Local Similarity 96.4%; Pred. No. 2.2e-11;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKTKRNTYRRPQDVKPPGGQIVG 28
Db 5 PKPQKTKRNTYRRPQDVKPPGGQIVG 32

RESULT 5
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 94.2%; Score 145; DB 10; Length 97;
Best Local Similarity 96.4%; Pred. No. 2.3e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1 PKPQRTKTRNTYRRPDVKFPGGGQIVG 28
||||| 18 PKPQRTKTRNTYRRPDVKFPGGGQIVG 32

RESULT 6
US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
; FILE OF INVENTION: applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: EP 00402225.7
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT

; ORGANISM: Hepatitis C virus
US-09-921-397-77
Query Match 94.2%; Score 145; DB 10; Length 103;
Best Local Similarity 96.4%; Pred. No. 2.5e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQRTKTRNTYRRPDVKFPGGGQIVG 28
Db 18 PKPQRTKTRNTYRRPDVKFPGGGQIVG 45
RESULT 7
US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGN
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95670076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-851-138-14

Query Match 94.2%; Score 145; DB 10; Length 108;
Best Local Similarity 96.4%; Pred. No. 2.6e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQRTKTRNTYRRPDVKFPGGGQIVG 28
Db 5 PKPQRTKTRNTYRRPDVKFPGGGQIVG 32
RESULT 8
US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS

;; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
;; TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
;; FILE OF INVENTION: applications thereof
;; FILE REFERENCE: B4809A - JAZ
;; CURRENT APPLICATION NUMBER: US/09/921.397
;; CURRENT FILING DATE: 2001-08-02
;; PRIOR APPLICATION NUMBER: EP 00402225.7
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 156
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 78
;; LENGTH: 113
;; TYPE: PRT
;; ORGANISM: Hepatitis C virus
US-09-921-397-78

Query Match 94.2%; Score 145; DB 10; Length 113;
Best Local Similarity 96.4%; Pred. No. 2.8e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 PKPQKTKRNTYRRPQDVKFGGGQIVG 28
|||||
5 PKPQKTKRNTNRRPQDVKFGGGQIVG 32

RESULT 9
US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851.138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836.075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 137 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:

US-09-851-138-46
Query Match 94.2%; Score 145; DB 10; Length 137;
Best Local Similarity 96.4%; Pred. No. 3.4e-13;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 PKPQKTKRNTYRRPQDVKFGGGQIVG 28
|||||
Db 5 PKPQKTKRNTNRRPQDVKFGGGQIVG 32
RESULT 10
US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNO:
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851.138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836.075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-851-138-60

Query Match 94.2%; Score 145; DB 10; Length 138;
Best Local Similarity 96.4%; Pred. No. 3.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKTKRNTYRRPQDVKFGGGQIVG 28
|||||
Db 5 PKPQKTKRNTNRRPQDVKFGGGQIVG 32

RESULT 11
US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy

```
US-09-899-046-42  
; Sequence 42, Application US/09899046  
; Publication No. US2003008274A1  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: New sequences of hepatitis C virus  
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.  
; NUMBER OF SEQUENCES: 270  
; COMPUTER READABLE FORM:
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-42

Query Match      94.2%; Score 145; DB 11; Length 169;
Est Local Similarity 96.4%; Pred. No. 4.3e-13;
atches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 PKPQRTKRNTYRRPQDVKFPGGQIVG 28
        ||||| ||||| ||||| ||||| |||||
DB      5 PKPQRTKRNTYRRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 12:01:12
Job time : 14.3636 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 9.54545 Seconds
(without alignments)
282.095 Million cell updates/sec

Title: US-09-491-146a-26

Perfect score: 156

Sequence: 1 PKPQRKPNTNRRPDVVFPGGGQIVG 28

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 283308 seqs, 9616862 residues
tal number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_76:*

1: pirl:*

2: pirl2:*

3: pirl3:*

4: pirl4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	142	91.0	108	2 S41353	genome polyprotein
2	142	91.0	108	2 S41355	genome polyprotein
3	142	91.0	108	2 S41357	genome polyprotein
4	142	91.0	108	2 S41348	genome polyprotein
5	142	91.0	112	2 S41371	genome polyprotein
6	142	91.0	112	2 S41341	genome polyprotein
7	142	91.0	114	2 S41370	genome polyprotein
8	142	91.0	114	2 S41369	genome polyprotein
9	142	91.0	114	2 S41368	genome polyprotein
10	142	91.0	115	2 S41342	genome polyprotein
11	142	91.0	115	2 S41344	genome polyprotein
12	142	91.0	115	2 S41350	genome polyprotein
13	142	91.0	115	2 S41354	genome polyprotein
14	142	91.0	115	2 S41347	genome polyprotein
15	142	91.0	115	2 S41345	genome polyprotein
16	142	91.0	118	2 S41343	genome polyprotein
17	142	91.0	118	2 S41346	genome polyprotein
18	142	91.0	369	2 S21471	genome polyprotein
19	142	91.0	441	2 S12707	genome polyprotein
20	142	91.0	513	2 PC1284	genome polyprotein
21	142	91.0	520	2 JQ1925	genome polyprotein
22	142	91.0	523	2 JQ1926	polyprotein - hepa
23	142	91.0	550	2 JH0711	genome polyprotein
24	142	91.0	782	2 S19876	genome polyprotein
25	142	91.0	782	2 S18031	genome polyprotein
26	142	91.0	782	2 S18032	genome polyprotein
27	142	91.0	787	2 PN0677	hypothetical prote
28	142	91.0	874	2 JQ0883	genome polyprotein
29	142	91.0	874	2 JQ0881	genome polyprotein

30 142 91.0 876 2 PC2219 polypeptide - hepa
31 142 91.0 3010 1 GNWVTC genome polyprotein
32 142 91.0 3010 1 GNWVCJ genome polyprotein
33 142 91.0 3010 1 S18030 genome polyprotein
34 142 91.0 3011 1 GNWVCH genome polyprotein
35 142 91.0 3011 1 S40770 genome polyprotein
36 142 91.0 3014 1 JC5620 genome polyprotein
37 142 91.0 3033 1 JQ1303 genome polyprotein
38 142 91.0 3033 1 GNWVJ8 genome polyprotein
39 138 88.5 115 2 S41351 genome polyprotein
40 138 88.5 115 2 S41349 genome polyprotein
41 138 88.5 125 2 S41352 genome polyprotein
42 138 88.5 640 2 JQ1584 genome polyprotein
43 138 88.5 3011 1 GNWVC3 genome polyprotein
44 137 87.8 88 2 S21336 genome polyprotein
45 137 87.8 109 2 S41367 genome polyprotein

ALIGNMENTS

RESULT 1

S41353
genome polyprotein - hepatitis C virus (genotype 2, N2) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N2
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41353
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41353
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:Z29456
A:Experimental source: genotype 2, N2
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 91.0%; Score 142; DB 2; Length 108;
Best Local Similarity 92.9%; Pred. No. 1.4e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQRKPNTNRRPDVVFPGGGQIVG 28
||||| |||||||||
Db 5 PKPQRKPNTNRRPDVVFPGGGQIVG 32

RESULT 2

S41355
genome polyprotein - hepatitis C virus (genotype 2, N4) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41355
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
A:Reference number: S41341
A:Accession: S41355
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:Z29458
A:Experimental source: genotype 2, N4
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 91.0%; Score 142; DB 2; Length 108;


```

Best Local Similarity 92.9%; Pred. No. 1.4e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 3
S41357
genome polyprotein - hepatitis C virus (genotype 2, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 2, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
Accession: S41357
Molecule type: genomic RNA
Residues: 1-108 <VAN>
Cross-references: EMBL:229460
Experimental source: genotype 2, N6
Superfamily: hepatitis C virus genome polyprotein
Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 91.0%; Score 142; DB 2; Length 108;
Best Local Similarity 92.9%; Pred. No. 1.4e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 4
S41348
genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N6
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
Accession: S41348
Molecule type: genomic RNA
Residues: 1-108 <VAN>
Cross-references: EMBL:229451
Experimental source: genotype 1, N6
Superfamily: hepatitis C virus genome polyprotein
Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Query Match 91.0%; Score 142; DB 2; Length 108;
Best Local Similarity 92.9%; Pred. No. 1.4e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 5
S41371
genome polyprotein - hepatitis C virus (genotype 5, N5) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus

```

```

A:Variety: genotype 5, N5
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S41371
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
Accession: S41371
Molecule type: genomic RNA
Residues: 1-112 <VAN>
Cross-references: EMBL:229474
Experimental source: genotype 5, N5
Superfamily: hepatitis C virus genome polyprotein
Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>

Query Match 91.0%; Score 142; DB 2; Length 112;
Best Local Similarity 92.9%; Pred. No. 1.5e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 6
S41341
genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 1, N1
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
Accession: S41341
Molecule type: genomic RNA
Residues: 1-112 <VAN>
Cross-references: EMBL:229444; NID:g443850; PIDN:CAA82582.1; PID:g443851
Experimental source: genotype 1, N1
Superfamily: hepatitis C virus genome polyprotein
Keywords: capsid protein; core protein; polyprotein
F:1-112/Product: core protein #status predicted <MAT>

Query Match 91.0%; Score 142; DB 2; Length 112;
Best Local Similarity 92.9%; Pred. No. 1.5e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 7
S41370
genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 5, N4
C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
submitted to the EMBL Data Library, January 1994
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
A:Reference number: S41341
Accession: S41370
Molecule type: genomic RNA
Residues: 1-114 <VAN>
Cross-references: EMBL:229473; NID:g443908; PIDN:CAA82611.1; PID:g443909
Experimental source: genotype 5, N4
Superfamily: hepatitis C virus genome polyprotein

```


QY 1 PKPQKPNRNTNRRPQDVKFPGGQIVG 28
||||| |||||||||

Result No.	Score	Query		ID	Description
		Match	Length		
1	142	91.0	513	POLG_HCVJ2	P27959 hepatitis c
2	142	91.0	520	POLG_HCVH4	Q01404 hepatitis c
3	142	91.0	520	POLG_HCVHK	Q01403 hepatitis c
4	142	91.0	737	POLG_HCVU5	P27960 hepatitis c
5	142	91.0	737	POLG_HCVJ7	P27961 hepatitis c
6	142	91.0	3010	POLG_HCVBK9	P26663 h genome po
7	142	91.0	3010	POLG_HCVJ3A	P26662 h genome po
8	142	91.0	3033	POLG_HCVH	P27958 h genome po
9	142	91.0	3033	POLG_HCVJ6	P26660 h genome po
10	142	91.0	3033	POLG_HCVJ8	P26661 h genome po
11	138	88.5	3011	POLG_HCVJ1	P26664 h genome po
12	135	86.5	3010	POLG_HCVTW	P29846 h genome po
13	134	85.9	3010	POLG_HCVJ7	Q00269 h genome po
14	55.5	35.6	586	2A5D_RABIT	Q28653 o serine/th
15	53.5	34.3	602	2A5D_HUMAN	Q14738 h serine/th
16	53	34.0	1311	FMR2_HUMAN	P51815 homo sapien
17	52	33.3	349	SCD6_YEAST	P45978 saccharomyc
18	51	32.7	167	YF57_ARCFU	O28715 archaeoglob
19	51	32.7	436	GDF6_BOVIN	P55106 bos taurus
20	50.5	32.4	303	YB1_XENLA	P21573 xenopus lae
21	50	32.1	282	RK4_TOBAC	O80361 nicotiana t
22	50	32.1	339	CSP_PLABE	P06915 plasmodium
23	50	32.1	347	CSP_PLABA	P23093 plasmodium
24	49	31.4	395	SCD_DROME	P49415 drosophila
25	49	31.4	1233	MAK4_MOUSE	P97820 mus musculu
26	49	31.4	1252	SC25_YEAST	P14771 saccharomyc
27	48	30.8	262	COAT_SHVX	Q04572 shallot vir
28	48	30.8	461	STR3_RAT	P58405 rattus norv
29	48	30.8	681	RP3A_MOUSE	P47708 mus musculu
30	48	30.8	584	RP3A_RAT	P47709 rattus norv
31	48	30.8	796	STR3_MOUSE	Q9692 mus musculu
32	48	30.8	993	POLG_TBDVP	P25220 avian infec
33	48	30.8	1001	P5PA_RAT	Q91mc1 rattus norv

CC	EMBL; D10074; BAA00968.1; -	DR
DR	InterPro: IPR002522; HCV_capsid.	DR
DR	InterPro: IPR002521; HCV_core.	DR
DR	InterPro: IPR002519; HCV_env.	DR
DR	InterPro: IPR002531; HCV_NSI.	DR
DR	Pfam: PF01543; HCV_capsid; 1.	DR
DR	Pfam: PF01342; HCV_core; 1.	DR
DR	Pfam: PF01539; HCV_env; 1.	DR
DR	Pfam: PF01560; HCV_NSI; 1.	DR
DR	ProDom: PD186062; HCV_NSI; 1.	DR
KW	Polypeptide; Glycoprotein; Coat protein; Envelope protein;	
KW	Transmembrane; Nonstructural protein.	
FT	INIT_MET 1	FT
FT	1	FT
FT	CHAIN 1 115	FT
FT	CHAIN 116 191	FT
FT	CHAIN 192 383	FT
FT	CHAIN 193 383	FT
FT	CHAIN 194 383	FT
FT	CHAIN 195 383	FT
FT	CHAIN 196 383	FT
FT	CHAIN 197 383	FT
FT	CHAIN 198 383	FT
FT	CHAIN 199 383	FT
FT	CHAIN 200 383	FT
FT	CHAIN 201 383	FT
FT	CHAIN 202 383	FT
FT	CHAIN 203 383	FT
FT	CHAIN 204 383	FT
FT	CHAIN 205 383	FT
FT	CHAIN 206 383	FT
FT	CHAIN 207 383	FT
FT	CHAIN 208 383	FT
FT	CHAIN 209 383	FT
FT	CHAIN 210 383	FT
FT	CHAIN 211 383	FT
FT	CHAIN 212 383	FT
FT	CHAIN 213 383	FT
FT	CHAIN 214 383	FT
FT	CHAIN 215 383	FT
FT	CHAIN 216 383	FT
FT	CHAIN 217 383	FT
FT	CHAIN 218 383	FT
FT	CHAIN 219 383	FT
FT	CHAIN 220 383	FT
FT	CHAIN 221 383	FT
FT	CHAIN 222 383	FT
FT	CHAIN 223 383	FT
FT	CHAIN 224 383	FT
FT	CHAIN 225 383	FT
FT	CHAIN 226 383	FT
FT	CHAIN 227 383	FT
FT	CHAIN 228 383	FT
FT	CHAIN 229 383	FT
FT	CHAIN 230 383	FT
FT	CHAIN 231 383	FT
FT	CHAIN 232 383	FT
FT	CHAIN 233 383	FT
FT	CHAIN 234 383	FT
FT	CHAIN 235 383	FT
FT	CHAIN 236 383	FT
FT	CHAIN 237 383	FT
FT	CHAIN 238 383	FT
FT	CHAIN 239 383	FT
FT	CHAIN 240 383	FT
FT	CHAIN 241 383	FT
FT	CHAIN 242 383	FT
FT	CHAIN 243 383	FT
FT	CHAIN 244 383	FT
FT	CHAIN 245 383	FT
FT	CHAIN 246 383	FT
FT	CHAIN 247 383	FT
FT	CHAIN 248 383	FT
FT	CHAIN 249 383	FT
FT	CHAIN 250 383	FT
FT	CHAIN 251 383	FT
FT	CHAIN 252 383	FT
FT	CHAIN 253 383	FT
FT	CHAIN 254 383	FT
FT	CHAIN 255 383	FT
FT	CHAIN 256 383	FT
FT	CHAIN 257 383	FT
FT	CHAIN 258 383	FT
FT	CHAIN 259 383	FT
FT	CHAIN 260 383	FT
FT	CHAIN 261 383	FT
FT	CHAIN 262 383	FT
FT	CHAIN 263 383	FT
FT	CHAIN 264 383	FT
FT	CHAIN 265 383	FT
FT	CHAIN 266 383	FT
FT	CHAIN 267 383	FT
FT	CHAIN 268 383	FT
FT	CHAIN 269 383	FT
FT	CHAIN 270 383	FT
FT	CHAIN 271 383	FT
FT	CHAIN 272 383	FT
FT	CHAIN 273 383	FT
FT	CHAIN 274 383	FT
FT	CHAIN 275 383	FT
FT	CHAIN 276 383	FT
FT	CHAIN 277 383	FT
FT	CHAIN 278 383	FT
FT	CHAIN 279 383	FT
FT	CHAIN 280 383	FT
FT	CHAIN 281 383	FT
FT	CHAIN 282 383	FT
FT	CHAIN 283 383	FT
FT	CHAIN 284 383	FT
FT	CHAIN 285 383	FT
FT	CHAIN 286 383	FT
FT	CHAIN 287 383	FT
FT	CHAIN 288 383	FT
FT	CHAIN 289 383	FT
FT	CHAIN 290 383	FT
FT	CHAIN 291 383	FT
FT	CHAIN 292 383	FT
FT	CHAIN 293 383	FT
FT	CHAIN 294 383	FT
FT	CHAIN 295 383	FT
FT	CHAIN 296 383	FT
FT	CHAIN 297 383	FT
FT		

FT CHAIN 384 >513 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 513 513
 SQ SEQUENCE 513 AA; 5570 MW; 943P31E3514CDEF3 CRC64;

Query Match 91.0%; Score 142; DB 1; Length 513;

Best Local Similarity 92.9%; Pred. No. 6.4e-13;

Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 PKPQKPNRNTNRPPQDVKFGGGQIVG 28

||||| ||||||||| ||||||||| |||||||

5 PKPQKTKRNTNRPPQDVKFGGGQIVG 32

RESULT 2

POLG_HCVH4 STANDARD; PRT; 520 AA.
 ID POLG_HCVH4
 AC Q01404;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 DE Hepatitis C virus (isolate HCV-476) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31643;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan."
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@sib-sib.ch.
 CC -----

DR EMBL; D10688; BAA01530.1;
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 >520 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT TRANSMEM 347 369 BY SIMILARITY.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Query Match 91.0%; Score 142; DB 1; Length 520;

Best Local Similarity 92.9%; Pred. No. 6.5e-13;

Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRPPQDVKFGGGQIVG 28

||||| ||||||||| ||||||||| |||||||

5 PKPQKTKRNTNRPPQDVKFGGGQIVG 32

RESULT 3

POLG_HCVHK STANDARD; PRT; 520 AA.
 ID POLG_HCVHK
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 DE Hepatitis C virus (isolate HCV-KF) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31644;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan."
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@sib-sib.ch.
 CC -----

DR EMBL; D10687; BAA01529.1;
 DR PIR; J01925; J01925.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.

FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN E (POTENTIAL).
 FT CHAIN 384 >520 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 347 369 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CARBOHYD 196 196 BY SIMILARITY.
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6F27349B CRC64;
 Query Match 91.0%; Score 142; DB 1; Length 520;
 Best Local Similarity 92.9%; Pred. No. 6.5e-13;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
 DB 5 PKPQKTRNTNRRPDQVKFPGGGQIVG 32
 RESULT 4
 POLG_HCVJ5 STANDARD; PRT; 737 AA.
 ID POLG_HCVJ5
 AC P27960.
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (Isolate HC-J5) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RA *full-length sequenced of a hepatitis C virus genome having poor
 RA homology to reported isolates: comparative study of four distinct
 RA genotypes*;
 RA Virology 188:331-341(1992).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D10075; BAA00969.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR Pfam: PF01543; HCV_capsid; 1.

DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN E (POTENTIAL).
 FT CHAIN 384 733 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 734 >737 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 737
 SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD50131 CRC64;
 Query Match 91.0%; Score 142; DB 1; Length 737;
 Best Local Similarity 92.9%; Pred. No. 9.5e-13;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
 DB 5 PKPQKTRNTNRRPDQVKFPGGGQIVG 32
 RESULT 5
 POLG_HCVJ7 STANDARD; PRT; 737 AA.
 ID POLG_HCVJ7
 AC P27961;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural
 DE proteins NS1 and NS2] (Fragment).
 OS Hepatitis C virus (Isolate HC-J7) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S.I., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RA *full-length sequenced of a hepatitis C virus genome having poor
 RA homology to reported isolates: comparative study of four distinct
 RA genotypes*;
 RA Virology 188:331-341(1992).
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D10075; BAA00969.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR Pfam: PF01543; HCV_capsid; 1.

DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 " Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 " Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 " 3D-structure.
 " INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 " CELLULAR AMINOPeptIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1039 1047 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1059 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1068 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1075 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1077 1081 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1082 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 1086 1087 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1088 1087 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1090 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1093 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1095 1097 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 1101 1103 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT TURN 1104 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 1108 1112
 FT STRAND 1120 1120
 FT STRAND 1122 1122
 FT STRAND 1129 1133
 FT TURN 1135 1136
 FT STRAND 1139 1144
 FT STRAND 1149 1157
 FT HELIX 1158 1161
 FT TURN 1162 1163
 FT STRAND 1165 1166
 FT STRAND 1168 1171
 FT TURN 1172 1174
 FT STRAND 1175 1186
 FT TURN 1187 1188
 FT STRAND 1189 1197
 FT HELIX 1198 1202
 FT TURN 1203 1204
 FT STRAND 1680 1688
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;
 Query Match 91.0%; Score 142; DB 1; Length 3010;
 Best Local Similarity 92.9%; Pred. No. 4.5e-12;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKPNRNTNRPPQDVKEPGGGQIVG 28
 DB 5 PKPQKTRNTNRPPQDVKEPGGGQIVG 32
 RESULT 7
 POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-9108550; PubMed-2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE-91192160; PubMed-1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
 RA Ohkoshi S., Shimotohno K.;
 RT "Molecular structure of the Japanese hepatitis C viral genome";
 RL FEBS Lett. 280:325-328(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA4534.1; -
 CC PIR; A36814; GNVVCH.
 CC PDB; 1HEI; 25-NOV-98.
 CC PDB; 1A1V; 16-FEB-99.
 CC PDB; 1A1B; 17-JUN-98.
 CC MEROPS; S29.001; -
 CC MEROPS; U39.001; -
 CC TRANSFAC; T04155; -
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002522; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_env.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4a.
 CC InterPro; IPR001490; HCV_NS4b.
 CC InterPro; IPR002868; HCV_NS5a.
 CC InterPro; IPR002166; HCV_RdRP.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVir.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.
 CC Pfam; PF01539; HCV_env; 1.
 CC Pfam; PF01560; HCV_NS1; 1.
 CC Pfam; PF01538; HCV_NS2; 1.
 CC Pfam; PF02907; HCV_NS3; 1.
 CC Pfam; PF01006; HCV_NS4a; 1.
 CC Pfam; PF01001; HCV_NS4b; 1.
 CC Pfam; PF01506; HCV_NS5a; 1.
 CC Pfam; PF00271; Helicase_C; 1.
 CC Pfam; PF00998; Viral_RdRP; 1.
 CC ProDom; PD186062; HCV_NS1; 1.
 CC SMART; SM00487; DEXdc; 1.
 CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 CC 3D-structure.
 CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC CHAIN 1 191
 CC CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.
 CC CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.
 CC CHAIN 747 809 PROTEIN P7.
 CC CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.
 CC CHAIN 1027 1657 PROTEASE/HELICASE NS3.
 CC CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.
 CC CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.
 CC CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
 CC CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
 CC CHAIN 3011 369 POTENTIAL.
 CC TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC NP_BIND 1230 1237 ATP (POTENTIAL).
 CC SITE 1316 1319 DECH BOX.

FT CARBOHYD 196
 FT CARBOHYD 209
 FT CARBOHYD 234
 FT CARBOHYD 305
 FT CARBOHYD 417
 FT CARBOHYD 423
 FT CARBOHYD 430
 FT CARBOHYD 448
 FT CARBOHYD 476
 FT CARBOHYD 532
 FT CARBOHYD 540
 FT CARBOHYD 556
 FT CARBOHYD 576
 FT CARBOHYD 623
 FT CARBOHYD 645
 FT STRAND 1224
 FT TURN 1232
 FT TURN 1233
 FT TURN 1236
 FT TURN 1239
 FT TURN 1246
 FT TURN 1247
 FT STRAND 1251
 FT HELIX 1258
 FT TURN 1272
 FT TURN 1271
 FT STRAND 1277
 FT TURN 1281
 FT STRAND 1282
 FT STRAND 1283
 FT STRAND 1291
 FT HELIX 1296
 FT TURN 1302
 FT STRAND 1312
 FT TURN 1317
 FT TURN 1323
 FT TURN 1336
 FT TURN 1343
 FT TURN 1352
 FT TURN 1360
 FT STRAND 1362
 FT STRAND 1368
 FT STRAND 1373
 FT TURN 1376
 FT STRAND 1378
 FT STRAND 1380
 FT HELIX 1382
 FT STRAND 1385
 FT STRAND 1389
 FT HELIX 1397
 FT TURN 1410
 FT STRAND 1414
 FT TURN 1419
 FT STRAND 1432
 FT TURN 1438
 FT STRAND 1439
 FT STRAND 1450
 FT STRAND 1456
 FT STRAND 1471
 FT STRAND 1480
 FT HELIX 1481
 FT TURN 1489
 FT STRAND 1497
 FT STRAND 1501
 FT STRAND 1507
 FT STRAND 1511
 FT HELIX 1514
 FT HELIX 1527
 FT STRAND 1532
 FT STRAND 1550
 FT HELIX 1555
 FT STRAND 1570
 FT TURN 1579
 FT STRAND 1580
 FT HELIX 1584
 FT TURN 1598
 FT HELIX 1606
 FT TURN 1614
 FT STRAND 1622
 FT STRAND 1627
 FT STRAND 1635
 FT HELIX 1640
 FT SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 91.0%; Score 142; DB 1; Length 3011;
Best Local Similarity 92.9%; Pred. No. 4.5e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQKPNRNTNRPPQDVKFPFGGQIVG 28
||||| ||||||| ||||||| |||||||

DB 5 PKPQKTKRNTNRPPQDVKFPFGGQIVG 32

RESULT 9
POLG_HCVJ6 STANDARD; PRT; 3033 AA.

AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Liuka H.,
RA Machida A., Miyakawa Y., Miyami M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: D00944; BAA00792.1; -
PIR: J01303; J01303.
DR HSP: P27958; 1HEI.
DR MEROPS: S29.001; -
DR MEROPS: U39.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR004109; HCV NS3.
DR InterPro: IPR000745; HCV NS4A.
DR InterPro: IPR001490; HCV NS4B.

DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSV1.
DR Pfam: PF01543; HCV_capsid; 1;
DR Pfam: PF01542; HCV_core; 1;
DR Pfam: PF01539; HCV_env; 1;
DR Pfam: PF01560; HCV_NS1; 1;
DR Pfam: PF01538; HCV_NS2; 1;
DR Pfam: PF02907; HCV_NS3; 1;
DR Pfam: PF01006; HCV_NS4a; 1;
DR Pfam: PF01001; HCV_NS4b; 1;
DR Pfam: PF01506; HCV_NS5a; 1;
DR Pfam: PF00271; Helicase_C; 1;
DR Pfam: PF00998; Viral_RdRP; 1;
DR ProDom: PD186062; HCV_NS1; 1;
DR SMART: SM00487; DEKDC; 1;
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Coat protein; Hydrolase; Serine protease;
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 91.0%; Score 142; DB 1; Length 3033;
Best Local Similarity 92.9%; Pred. No. 4.6e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQKPNRNTNRPPQDVKFPFGGQIVG 28
||||| ||||||| ||||||| |||||||

DB 5 PKPQKTKRNTNRPPQDVKFPFGGQIVG 32

RESULT 10
POLG_HCVJ8 STANDARD; PRT; 3033 AA.
ID POLG_HCVJ8
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: M62321; AAA45676.1; -
 CC PIR: A39166; GNWC3.
 CC PDB: 1A1V; 16-FEB-99.
 CC PDB: 1HEI; 25-NOV-98.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RDRP.
 CC InterPro: IPR001650; Helicase.C.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00271; helicase.C; 1.
 CC Pfam: PF00998; Viral_RDRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXdc; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 POTENTIAL.
 FT ACT_SITE 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 ATP (POTENTIAL).
 FT SITE 1316 DECH BOX.
 FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 88.5%; Score 138; DB 1; Length 3011;
 Best Local Similarity 89.3%; Pred. No. 1.7e-11;

Matches 25; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKQKKNRNTNRRPQDVKFGGQIVG 28
 DB 5 PKQKKNRNTNRRPQDVKFGGQIVG 32

RESULT 12

POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P2);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID:31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:92230206; PubMed:1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84754; .; NOT_ANNOTATED_CDS.
 DR PIR: A40244; GNVTVW
 DR PDB; 1N64; 25-FEB-03.
 DR PDB; INS3; 08-APR-98.
 DR MEROPS; S29.001; .
 DR MEROPS; U39.001; .
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 KW INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423

FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
 Query Match 86.5%; Score 135; DB 1; Length 3010;
 Best Local Similarity 92.6%; Pred. No. 4.6e-11;
 Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 KPQRKPNRTNRRPDVKFPGGGQIVG 28
 DB 6 KPQRKTRNTNRRPDVKFPGGGQIVG 32
 RESULT 13
 POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

[illegible]

SQ SEQUENCE 586 AA; 68090 MW; EI49A309CDDA7495 CRC64;
 Query Match 35.6%; Score 55.5; DB 1; Length 586;
 Best Local Similarity 35.1%; Pred. No. 2.2;
 Matches 13; Conservative 8; Mismatches 5; Indels 11; Gaps 2;
 QY 1 PKQKRNRTNRPPQD-----VKPPGGGQIV 27
 Db 30 PQPOSQP-PSNKRPSNSTPPTQLSKIRYGGQIV 65
 RESULT 15
 2A5D_HUMAN
 ID 2A5D_HUMAN STANDARD; PRT; 602 AA.
 AC Q14738; Q00494; Q00696; Q15171;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Serine/threonine protein phosphatase 2A, 56 kDa regulatory subunit,
 B56 delta isoform (PP2A, B subunit, B' delta isoform) (PP2A, B subunit,
 subunit, R5 delta isoform).
 GN PPP2R5D.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM DELTA-1).
 RC TISSUE=Fetal brain;
 RX MEDLINE=96355607; PubMed=8703017;
 RT McCright B., Rivers A.M., Audlin S., Virshup D.M.;
 RT "The B56 family of protein phosphatase 2A (PP2A) regulatory subunits
 RT encodes differentiation-induced phosphoproteins that target PP2A to
 RT both nucleus and cytoplasm".
 RL J. Biol. Chem. 271:22081-22089(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS DELTA-1 AND DELTA-3).
 RC TISSUE=Brain cortex;
 RX MEDLINE=97324098; PubMed=9180267;
 RT Tanabe O., Gomez G.A., Nishito Y., Usui H., Takeda M.;
 RT "Molecular heterogeneity of the cDNA encoding a 74-kDa regulatory
 RT subunit (B' or delta) of human protein phosphatase 2A".
 RL FEBS Lett. 408:52-56(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM DELTA-2), AND SEQUENCE OF 501-508;
 RP 550-559; 573-580 AND 584-601 (DELTA-1).
 RC TISSUE=Bone marrow, and Brain cortex;
 RX MEDLINE=96159032; PubMed=8566219;
 RT Tanabe O., Nagase T., Murakami T., Nozaki H., Usui H., Nishito Y.,
 RT Hayashi H., Kagamiyama H., Takeda M.;
 RT "Molecular cloning of a 74-kDa regulatory subunit (B' or delta) of
 RT human protein phosphatase 2A".
 RL FEBS Lett. 379:107-111(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS DELTA-1 AND DELTA-2).
 RC TISSUE=Colon, Eye, and Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RT Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
 RA Villalón D., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Whiting J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
 CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
 CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
 CC COMPARTMENT.
 CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
 CC COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
 CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES
 CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
 CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS
 CC B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56
 CC FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
 CC AND CELL SIGNALING MOLECULES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR IN INTERPHASE,
 CC NUCLEAR DURING MITOSIS.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Name-Delta-1;
 CC IsoId=Q14738-1; Sequence-Displayed;
 CC Name-Delta-2;
 CC IsoId=Q14738-2; Sequence-VSP_005111;
 CC Name-Delta-3;
 CC IsoId=Q14738-3; Sequence-VSP_005110;
 CC -1- TISSUE SPECIFICITY: ISOFORM DELTA-2 IS WIDELY EXPRESSED. ISOFORM
 CC DELTA-1 IS HIGHLY EXPRESSED IN BRAIN.
 CC -1- INDUCTION: BY RETINOIC ACID; IN NEUROBLASTOMA CELL LINES.
 CC -1- PTM: AT LEAST ISOFORM DELTA-1 IS PHOSPHORYLATED ON SERINE
 CC RESIDUES.
 CC -1- SIMILARITY: Belongs to the phosphatase 2A regulatory subunit B
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L76702; AAB69751.1; -;
 DR EMBL; AB000634; BAA20381.1; -;
 DR EMBL; AB000635; BAA20382.1; -;
 DR EMBL; D78360; BAA11372.1; -;
 DR EMBL; BC001095; AAH01095.1; -;
 DR EMBL; BC001175; AAH01175.1; -;
 DR EMBL; BC010692; AAH10692.1; -;
 DR PIR; S68686; S68686.
 DR Genbank; HGNC:9312; PPP2R5D.
 DR GK; Q14738; -;
 DR MM; 601646; -;
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0008601; F:protein phosphatase type 2A, regulator acti. .; TAS.
 DR GO; GO:0007399; F:neurogenesis; TAS.
 DR InterPro; IPR002554; B56.
 DR Pfam; PF01603; B56; 1.
 KW Nuclear protein; Phosphorylation; Alternative splicing; Repeat;
 KW Domain family.
 FT DOMAIN 37 52 8 X 2 AA APPROXIMATE TANDEM REPEATS OF Q-
 FT P.
 FT DOMAIN 523 530 SH3 BINDING, CLASS I (POTENTIAL).
 FT DOMAIN 548 565 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT VARSPLIC 11 116 Missing (in isoform Delta-3).
 FT VARSPLIC 85 116 Missing (in isoform Delta-2).
 FT VARSPLIC 602 AA; 69991 MW; F15F71AF4E565307 CRC64;
 FT SEQUENCE 34.3%; Score 53.5; DB 1; Length 602;
 Query Match 34.3%; Score 53.5; DB 1; Length 602;
 Best Local Similarity 31.0%; Pred. No. 4.4;

Qy 1 PKPQRKPNR-----NTNRRRPQD-----VAFPPGGGQIV 27
 |::| : | ::||| : :| : |||
Dd 40 PGPQPQPAQS QPPSSNKRRFSNSTPTTQLSKIRYSGGPQIV 81

Search completed: August 7, 2003, 11:19:59
Job time : 5.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:05:41 ; Search time 25.6364 Seconds
(without alignments)
281.845 Million cell updates/sec

Title: US-09-491-146A-26

Perfect score: 156
Sequence: 1 PKPQKPNRNTNRRPDQVKFPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriopl.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	96.2	191	12 Q8V7S2	Q8V7S2 hepatitis c
2	150	96.2	191	12 Q8V7T0	Q8V7T0 hepatitis c
3	150	96.2	191	12 Q8V7S4	Q8V7S4 hepatitis c
4	148	94.9	119	12 Q81291	Q81291 hepatitis c
5	148	94.9	191	12 Q8JP76	Q8JP76 hepatitis c
6	148	94.9	191	12 Q8V7X8	Q8V7X8 hepatitis c
7	148	94.9	191	12 Q8V7V4	Q8V7V4 hepatitis c
8	148	94.9	191	12 Q8V7X5	Q8V7X5 hepatitis c
9	148	94.9	191	12 Q8V7Q3	Q8V7Q3 hepatitis c
10	148	94.9	191	12 Q8V7K4	Q8V7K4 hepatitis c
11	148	94.9	191	12 Q8V7M4	Q8V7M4 hepatitis c
12	148	94.9	191	12 Q8V7L1	Q8V7L1 hepatitis c
13	148	94.9	191	12 Q8V7Y3	Q8V7Y3 hepatitis c
14	148	94.9	191	12 Q8V7N6	Q8V7N6 hepatitis c
15	148	94.9	191	12 Q8V7Y0	Q8V7Y0 hepatitis c
16	148	94.9	191	12 Q8V7K9	Q8V7K9 hepatitis c

17	148	94.9	191	12 Q8V7W1	Q8V7W1 hepatitis c
18	148	94.9	191	12 Q8V7W6	Q8V7W6 hepatitis c
19	148	94.9	191	12 Q8V7N7	Q8V7N7 hepatitis c
20	148	94.9	191	12 Q8V7X7	Q8V7X7 hepatitis c
21	148	94.9	191	12 Q8V7V9	Q8V7V9 hepatitis c
22	148	94.9	191	12 Q8V7L6	Q8V7L6 hepatitis c
23	148	94.9	191	12 Q8V7M0	Q8V7M0 hepatitis c
24	148	94.9	191	12 Q8V7K8	Q8V7K8 hepatitis c
25	148	94.9	191	12 Q8JW46	Q8JW46 hepatitis c
26	148	94.9	191	12 Q8V7L7	Q8V7L7 hepatitis c
27	148	94.9	191	12 Q8V7M7	Q8V7M7 hepatitis c
28	145	92.9	191	12 Q8V7V3	Q8V7V3 hepatitis c
29	145	92.9	191	12 Q8V7W0	Q8V7W0 hepatitis c
30	145	92.9	191	12 Q8V7W8	Q8V7W8 hepatitis c
31	145	92.9	191	12 Q8V7W5	Q8V7W5 hepatitis c
32	144	92.3	191	12 Q8V7Y2	Q8V7Y2 hepatitis c
33	144	92.3	802	12 Q68520	Q68520 hepatitis c
34	142	91.0	45	12 Q68311	Q68311 hepatitis c
35	142	91.0	45	12 Q68306	Q68306 hepatitis c
36	142	91.0	46	12 Q68309	Q68309 hepatitis c
37	142	91.0	60	12 Q8JYR9	Q8JYR9 hepatitis c
38	142	91.0	61	12 Q8JYR8	Q8JYR8 hepatitis c
39	142	91.0	61	12 Q8JYR5	Q8JYR5 hepatitis c
40	142	91.0	61	12 Q8JYR7	Q8JYR7 hepatitis c
41	142	91.0	62	12 Q8JYR6	Q8JYR6 hepatitis c
42	142	91.0	74	12 Q68712	Q68712 hepatitis c
43	142	91.0	74	12 Q68682	Q68682 hepatitis c
44	142	91.0	74	12 Q68708	Q68708 hepatitis c
45	142	91.0	74	12 Q68679	Q68679 hepatitis c

ALIGNMENTS

RESULT 1

Q8V7S2
ID Q8V7S2 PRELIMINARY; PRT; 191 AA.
AC Q8V7S2;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic lesions";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AB062202; BAB83396.1; -;
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 191 AA; 20740 MW; 64151CA83695EF34 CRC64;

Query Match 96.2%; Score 150; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 7e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQKPNRNTNRRPDQVKFPGGQIVG 28
||||| |

```
Db 5 PKPQKPKRNTNRRPDQVKFPGGGQIVG 32
RESULT 2
Q8V7T0
ID Q8V7T0 PRELIMINARY; PRT; 191 AA.
AC Q8V7T0;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
DE CORE.
GN Hepatitis C virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RA "Hepatitis C virus quasisppecies in cancerous and non-cancerous hepatic
lesions.";
RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AB062194; BAB83388.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_core.
DR Pfam; PF01543; HCV_core; 1.
DR Polyprotein. 191
KW NON_TER 191
SQ SEQUENCE 191 AA; 20740 MW; 641510CA83695EF34 CRC64;

Query Match 96.2%; Score 150; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 7e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKPKRNTNRRPDQVKFPGGGQIVG 28
||||| |||||||||
Db 5 PKPQKPKRNTNRRPDQVKFPGGGQIVG 32

RESULT 3
Q8V7S4
ID Q8V7S4 PRELIMINARY; PRT; 191 AA.
AC Q8V7S4;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
DE CORE.
GN Hepatitis C virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RA "Hepatitis C virus quasisppecies in cancerous and non-cancerous hepatic
lesions.";
RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AB062200; BAB83394.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; IPR002521; HCV_core.
DR Pfan; PF01543; HCV_capsid; 1.
```

```
DR Pfan; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191
SQ SEQUENCE 191 AA; 20744 MW; 640D7CA82765EF34 CRC64;

Query Match 96.2%; Score 150; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 7e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKPKRNTNRRPDQVKFPGGGQIVG 28
||||| |||||||||
Db 5 PKPQKPKRNTNRRPDQVKFPGGGQIVG 32

RESULT 4
Q81291
ID Q81291 PRELIMINARY; PRT; 119 AA.
AC Q81291;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Genome polyprotein (Fragment).
DE OS Hepatitis C virus.
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Td-34/92;
RX MEDLINE=9518942; PubMed=7883898;
RA Hotta H., Handajani R., Ingelusida M., Soemarto W., Doi H.,
RA Miyajima H., Homma M.;
RT "Subtype analysis of hepatitis C virus in Indonesia on the basis of
RT NS5b region sequences.";
RL J. Clin. Microbiol. 32:3049-3051(1994).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D26383; BAA05399.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfan; PF01543; HCV_capsid; 1.
KW Polyprotein. 119
FT NON_TER 119
SQ SEQUENCE 119 AA; F6836FBFA5E5EB2A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 119;
Best Local Similarity 96.4%; Pred. No. 8.2e-14;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQKPKRNTNRRPDQVKFPGGGQIVG 28
||||| |||||||||
Db 5 PKPQKPKRNTNRRPDQVKFPGGGQIVG 32

RESULT 5
Q8JP76
ID Q8JP76 PRELIMINARY; PRT; 191 AA.
AC Q8JP76;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Core protein (Genome polyprotein).
DE OS Hepatitis C virus.
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cored;
RA Polyak S.J., Miller K., McArdle S., Gale M.J. Jr., Tenover B.,
RA Hiscott J., Gretch D.R.;
RT "Quasispecies Dependent Activation of the Innate Interferon System By
```


RESULT 9

Q8V7Q3 ID Q8V7Q3 PRELIMINARY; PRT; 191 AA.
AC Q8V7Q3;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic lesions";
PI Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AB062221; BAB83415.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 191 AA; 20794 MW; 36DD3CBAE53B8532 CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRKPNNTNRRPDVKFPGGGQIVG 28
||||| ||||||| ||||||| |||||||
DB 5 PKPQRKTNRTNRRPDVKFPGGGQIVG 32

RESULT 10

Q8V7K4 ID Q8V7K4 PRELIMINARY; PRT; 191 AA.
AC Q8V7K4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic lesions";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AB062270; BAB83464.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 191 AA; 20794 MW; 36DD3CBAE53B8532 CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

SQ SEQUENCE 191 AA; 20718 MW; 23C8A9ACE83B849A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPQRKPNNTNRRPDVKFPGGGQIVG 28
||||| ||||||| ||||||| |||||||
DB 5 PKPQRKTNRTNRRPDVKFPGGGQIVG 32

RESULT 11

Q8V7M4 ID Q8V7M4 PRELIMINARY; PRT; 191 AA.
AC Q8V7M4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic lesions";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: AB062250; BAB83444.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER
SQ SEQUENCE 191 AA; 20746 MW; 23DB0B26D33B849A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPQRKPNNTNRRPDVKFPGGGQIVG 28
||||| ||||||| ||||||| |||||||
DB 5 PKPQRKTNRTNRRPDVKFPGGGQIVG 32

RESULT 12

Q8V7L1 ID Q8V7L1 PRELIMINARY; PRT; 191 AA.
AC Q8V7L1;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kato N.;
RT "Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic lesions";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AB062263; BAB83457.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191
SQ SEQUENCE 191 AA; 20730 MW; 23C8A9ACE82A749A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPDQVKGPGGQIVG 28
Db 5 PKPQKTNRTNRRPDQVKGPGGQIVG 32

RESULT 13
Q8V7Y3 PRELIMINARY; PRT; 191 AA.
ID Q8V7Y3
AC Q8V7Y3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN ID Q8V7Y0 PRELIMINARY; PRT; 191 AA.
RC TISSUE=Liver;
RA Kato N.;
RT Hepatitis C virus quasiespecies in cancerous and non-cancerous
RT lesions.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AB061996; BAB83335.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191
SQ SEQUENCE 191 AA; 20718 MW; 23C8A9ACE83B849A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPDQVKGPGGQIVG 28
Db 5 PKPQKTNRTNRRPDQVKGPGGQIVG 32

RESULT 14
Q8V7N6 PRELIMINARY; PRT; 191 AA.
ID Q8V7N6
AC Q8V7N6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
GN CORE.
OS Hepatitis C virus.

```

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11103;
RN ID Q8V7Y0 PRELIMINARY; PRT; 191 AA.
RC TISSUE=Liver;
RA Kato N.;
RT Hepatitis C virus quasiespecies in cancerous and non-cancerous hepatic
RT lesions.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AB062238; BAB83432.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191
SQ SEQUENCE 191 AA; 20730 MW; 23C8A9ACE82A749A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPDQVKGPGGQIVG 28
Db 5 PKPQKTNRTNRRPDQVKGPGGQIVG 32

RESULT 15
Q8V7Y0 PRELIMINARY; PRT; 191 AA.
ID Q8V7Y0
AC Q8V7Y0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Core protein (Genome polyprotein) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11103;
RN ID Q8V7Y0 PRELIMINARY; PRT; 191 AA.
RC TISSUE=Liver;
RA Kato N.;
RT Hepatitis C virus quasiespecies in cancerous and non-cancerous
RT lesions.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AB061999; BAB83338.1; -.
DR InterPro; IPR002522; HCV_capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
KW Polyprotein.
FT NON_TER 191
SQ SEQUENCE 191 AA; 20718 MW; 23C8A9ACE83B849A CRC64;

Query Match 94.9%; Score 148; DB 12; Length 191;
Best Local Similarity 96.4%; Pred. No. 1.4e-13;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPDQVKGPGGQIVG 28
Db 5 PKPQKTNRTNRRPDQVKGPGGQIVG 32

```

Fri Aug 8 09:20:10 2003

us-09-491-146a-26.rspt

Page 6

Search completed: August 7, 2003, 11:18:59
Job time : 26.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2003, 11:07:41 ; Search time 10.5455 Seconds
(without alignments)
112.343 Million cell updates/sec

Title: US-09-491-146A-26

Perfect score: 156
Sequence: 1 PPKQKPNRNTNRRPDVKFPGGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues
Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	28	3	US-08-921-887-26
2	142	91.0	34	3	US-08-380-160-6
3	142	91.0	43	4	US-09-020-846-36
4	142	91.0	44	3	US-08-380-160-2
5	142	91.0	44	4	US-09-389-756-1
6	142	91.0	45	3	US-08-380-160-1
7	142	91.0	61	1	US-07-946-054-9
8	142	91.0	61	1	US-08-083-947-23
9	142	91.0	61	1	US-08-530-550-3
10	142	91.0	61	1	US-08-462-037-26
11	142	91.0	61	5	PCT-US93-08638-9
12	142	91.0	61	5	PCT-US94-07088-23
13	142	91.0	61	5	PCT-US95-13660-3
14	142	91.0	74	3	US-08-836-075A-14
15	142	91.0	74	4	US-08-836-075A-108
16	142	91.0	100	4	US-08-635-886C-198
17	142	91.0	108	3	US-08-836-075A-14
18	142	91.0	115	1	US-08-324-977-8
19	142	91.0	115	2	US-08-384-616-8
20	142	91.0	115	2	US-08-904-686A-8
21	142	91.0	115	3	US-09-315-850-8
22	142	91.0	123	2	US-08-501-195-2
23	142	91.0	137	3	US-08-836-075A-46
24	142	91.0	138	3	US-08-836-075A-60
25	142	91.0	140	2	US-08-501-195-1
26	142	91.0	154	3	US-08-854-531-2
27	142	91.0	154	5	PCT-US95-13552-2

Query Match 100.0%; Score 156; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ALIGNMENTS

RESULT 1
US-08-921-887-26
; Sequence 26, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUDYAKOV, YURI E.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; METHOD OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SA
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-26

Sequence 16, Appl
Sequence 1, Appl1
Sequence 155, App
Sequence 156, App
Sequence 157, App
Sequence 158, App
Sequence 159, App
Sequence 160, App
Sequence 161, App
Sequence 163, App
Sequence 164, App
Sequence 165, App
Sequence 166, App
Sequence 167, App
Sequence 168, App
Sequence 169, App
Sequence 170, App
Sequence 171, App

OY 1 PKPQKPNRNTNRRPDVKFPGGGQIVG 28
 |||||
 DB 1 PKPQKPNRNTNRRPDVKFPGGGQIVG 28

RESULT 2

US-08-380-160-6
 ; Sequence 6, Application US/08380160
 ; Patent No. 6235284
 ; GENERAL INFORMATION:
 ; APPLICANT: DALBON, Pascal
 ; APPLICANT: JOLIVET, Michel
 ; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
 ; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OLIFF & BERRIDGE
 ; STREET: P.O. Box 19928
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22320
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/380,160
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/057,471
 ; FILING DATE: 06-MAY-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berridge, William P.
 ; REGISTRATION NUMBER: 30,024
 ; REFERENCE/DOCKET NUMBER: WPB 28682
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703)836-6400
 ; TELEFAX: (703)836-2787
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 34 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Human Hepatitis C Virus
 ; US-08-380-160-6

Query Match 91.0%; Score 142; DB 3; Length 34;
 Best Local Similarity 92.9%; Pred. No. 9.9e-13;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQKPNRNTNRRPDVKFPGGGQIVG 28
 |||||
 DB 4 PKPQKPNRNTNRRPDVKFPGGGQIVG 31

RESULT 3

US-09-020-846-36
 ; Sequence 36, Application US/09020846
 ; Patent No. 6322965
 ; GENERAL INFORMATION:
 ; APPLICANT: YAMAGUCHI, Kenjiro

APPLICANT: KASHIWAKUMA, Tomiko
 APPLICANT: CHIBA, Yukie
 APPLICANT: YAGI, Shintaro
 APPLICANT: HASEGAWA, Akira
 TITLE OF INVENTION: CHIMERA ANTIGEN PEPTIDE
 NUMBER OF SEQUENCES: 72
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY & LARDNER
 STREET: 3000 K Street, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20007-5109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/020,846
 FILING DATE: 09-FEB-1998
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 9-027015
 FILING DATE: 10-FEB-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-024045
 FILING DATE: 09-FEB-1996

ATTORNEY/AGENT INFORMATION:
 NAME: Wegner, Harold C.
 REGISTRATION NUMBER: 25,258
 REFERENCE/DOCKET NUMBER: 053466/0225
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 43 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear

MOLECULE TYPE: peptide
 US-09-020-846-36

Query Match 91.0%; Score 142; DB 4; Length 43;
 Best Local Similarity 92.9%; Pred. No. 1.3e-12;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKPQKPNRNTNRRPDVKFPGGGQIVG 28
 |||||
 DB 5 PKPQKPNRNTNRRPDVKFPGGGQIVG 32

RESULT 4

US-08-380-160-2
 ; Sequence 2, Application US/08380160
 ; Patent No. 6235284
 ; GENERAL INFORMATION:
 ; APPLICANT: DALBON, Pascal
 ; APPLICANT: JOLIVET, Michel
 ; TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
 ; TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESPECIALLY
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OLIFF & BERRIDGE
 ; STREET: P.O. Box 19928
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22320
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/380.160
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,471
FILING DATE: 06-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 28682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-6400
TELEFAX: (703)836-2787
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Human Hepatitis C Virus
STRAIN: H77
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..44
OTHER INFORMATION: /note= "N-terminal sequence of the
protein of the nucleocapside or CORE protein of
the human hepatitis C virus"
US-08-380-160-2

Query Match 91.0%; Score 142; DB 3; Length 44;
Best Local Similarity 92.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 PKPQKPNRNTNRRPQDVKFPGGGQIVG 28
||||| |||||||||
Db 4 PKPQKTKRNTNRRPQDVKFPGGGQIVG 31
ULT 5
-09-389-756-1
Sequence 1, Application US/09389756
Patent No. 6576240
GENERAL INFORMATION:
APPLICANT: JOLIVET, MICHEL
APPLICANT: PENIN, FRANCOIS
APPLICANT: DALBON, PASCAL
APPLICANT: LADAVIERE, LAURENT
APPLICANT: LACOUX, XAVIER
TITLE OF INVENTION: ANTIGENIC STRUCTURAL PEPTIDE, ANTIGENIC AND IMMUNOGENIC
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Human Hepatitis C Virus
STRAIN: H77
FEATURE:
NAME/KEY: Peptide
PUBLICATION INFORMATION:

AUTHORS: Ogata, N. et al.
TITLE: Nucleotide Sequence and Mutation Rate of the H Strain
Patent No. 6576240
TITLE: of Hepatitis Virus
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 3392-3396
DATE: 1991
RELEVANT RESIDUES: 2 TO 45
US-09-389-756-1
Query Match 91.0%; Score 142; DB 4; Length 44;
Best Local Similarity 92.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 PKPQKPNRNTNRRPQDVKFPGGGQIVG 28
||||| |||||||||
Db 4 PKPQKTKRNTNRRPQDVKFPGGGQIVG 31
TELEX:

RESULT 6
US-08-380-160-1
Sequence 1, Application US/08380160
Patent No. 6235284
GENERAL INFORMATION:
APPLICANT: DALBON, Pascal
APPLICANT: JOLIVET, Michel
TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
HEPATITIS C VIRUS (HCV), AND WHICH CAN BE USED ESPECIALI
TITLE OF INVENTION: FOR DETECTING THE LATTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: P.O. Box 19928
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/380.160
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,471
FILING DATE: 06-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 28682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-6400
TELEFAX: (703)836-2787
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Human Hepatitis C Virus
STRAIN: H77
FEATURE:
NAME/KEY: Peptide

```
;
; LOCATION: 1..45
; OTHER INFORMATION: /note= "N-terminal sequence of the
; OTHER INFORMATION: protein of the nucleocapsid or CORE protein of
; OTHER INFORMATION: the human hepatitis C virus"
US-08-380-160-1

Query Match          91.0%; Score 142; DB 3; Length 45;
Best Local Similarity 92.9%; Pred. No. 1.3e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPQDVKFPFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTKRNTNRRPQDVKFPFGGQIVG 32

RESULT 7
US-07-946-054-9
; Sequence 9, Application US/07946054
; Patent No. 5582968
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara H
; TITLE OF INVENTION: No. 5582968el Branched Hybrid and Cluster
; TITLE OF INVENTION: Peptides Effective in Diagnosing and Detecting
; TITLE OF INVENTION: No. 5582968-B Hepatitis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: United Biomedical Inc.
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: New York
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/946,054
; FILING DATE: 15-SEP-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 2000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-1717
; TELEFAX: 516-273-2828
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-946-054-9

Query Match          91.0%; Score 142; DB 1; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPQDVKFPFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 4 PKPQKTKRNTNRRPQDVKFPFGGQIVG 31

RESULT 8
US-08-083-947-23
; Sequence 23, Application US/08083947
; Patent No. 5639594
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
```

```
;
; TITLE OF INVENTION: No. 5639594el Linear and Branched Peptides Effective
; TITLE OF INVENTION: In Diagnosing and Detecting No. 5639594-A, No. 5639594-I
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,947
; FILING DATE: 19930628
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 946,054
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34045
; REFERENCE/DOCKET NUMBER: 2000Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)273-2828
; TELEFAX: (516)273-1717
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-083-947-23

Query Match          91.0%; Score 142; DB 1; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPQDVKFPFGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 4 PKPQKTKRNTNRRPQDVKFPFGGQIVG 31

RESULT 9
US-08-530-550-3
; Sequence 3, Application US/08530550
; Patent No. 5736321
; GENERAL INFORMATION:
; APPLICANT: Hosein, Barbara
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Peptides Effective for Diagnosis and
; TITLE OF INVENTION: Detection of Hepatitis C Infection
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: M. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: NY
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,550
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
```

NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34,045
REFERENCE/DOCKET NUMBER: 20002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-530-550-3

Query Match 91.0%; Score 142; DB 1; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
||||| ||||||||| ||||||||| |||||||
Db 4 PKPQKTKRNTNRRPDQVKFPGGGQIVG 31

RESULT 11
PCT-US93-08638-9
; Sequence 9, Application PC/TUS9308638
; GENERAL INFORMATION:
; APPLICANT: United Biomedical Inc.
; TITLE OF INVENTION: Novel Branched Hybrid and Cluster Peptides
; TITLE OF INVENTION: Effective in Diagnosing and Detecting Non-A,
; TITLE OF INVENTION: Non-B Hepatitis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: UNITED BIOMEDICAL INC.
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: New York
; COUNTRY: USA
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08638
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Lisa Wilson
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 9055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-2828
; TELEFAX: 516-273-1717
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08638-9

Query Match 91.0%; Score 142; DB 5; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
||||| ||||||||| ||||||||| |||||||
Db 4 PKPQKTKRNTNRRPDQVKFPGGGQIVG 31

RESULT 12
PCT-US94-07088-23
; Sequence 23, Application PC/TUS9407088
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Hosein, Barbara
; TITLE OF INVENTION: Novel Linear And Branched
; TITLE OF INVENTION: Peptides Effective in
; TITLE OF INVENTION: Diagnosing And Detecting
; TITLE OF INVENTION: Non-A, Non-B Hepatitis
; NUMBER OF SEQUENCES: 23

NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34,045
REFERENCE/DOCKET NUMBER: 20002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-530-550-3

Query Match 91.0%; Score 142; DB 1; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
||||| ||||||||| ||||||||| |||||||
Db 4 PKPQKTKRNTNRRPDQVKFPGGGQIVG 31

RESULT 10
US-08-262-037-26
; Sequence 26, Application US/08262037
; Patent No. 5747239
; GENERAL INFORMATION:
; APPLICANT: Chang Yi Wang and Barbara Hosein
; TITLE OF INVENTION: SYNTHETIC PEPTIDES SPECIFIC FOR
; TITLE OF INVENTION: THE DETECTION OF ANTIBODIES TO HCV, DIAGNOSIS OF HCV
; TITLE OF INVENTION: INFECTION AND PREVENTION THEREOF AS VACCINES
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/262,037
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/719,819
; FILING DATE: 24-June-1991
; APPLICATION NUMBER: 07/667,275
; FILING DATE: 11-Mar-1991
; APPLICATION NUMBER: 07/651,735
; FILING DATE: 07-Feb-1991
; APPLICATION NUMBER: 07/558,799
; FILING DATE: 26-July-1990
; APPLICATION NUMBER: 07/510,153
; FILING DATE: 16-April-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4043 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino acid
; STRANDEDNESS:

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Maria C.H. Lin
;; STREET: 345 Park Avenue
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10154
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/07088
;; FILING DATE: 22-JUNE-1994
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/083,947
;; FILING DATE: 28-JUNE-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: LIN, MARIA C.H.
;; REGISTRATION NUMBER: 29323
;; REFERENCE/DOCKET NUMBER: 1151-4101PC1
;; TELEPHONE: (212)758-4800
;; TELEFAX: (212)751-6849
;; INFORMATION FOR SEQ ID NO: 23:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 61 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
PCT-US94-07088-23

Query Match 91.0%; Score 142; DB 5; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPDQVKGPGGQIVG 28
||||| |||||||||
Db 4 PKPQKTKRNTNRRPDQVKGPGGQIVG 31

RESULT 13
PCT-US95-13660-3
;; Sequence 3, Application PC/TUS9513660
;; GENERAL INFORMATION:
;; APPLICANT: Hosein, Barbara
;; TITLE OF INVENTION: Peptides Effective for
;; TITLE OF INVENTION: Diagnosis and Detection of Hepatitis C Infection
;; NUMBER OF SEQUENCES: 34
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Morgan & Finnegan, L.L.P.
;; STREET: 345 Park Avenue
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10154
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version
;; SOFTWARE: #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/13660
;; FILING DATE: 23 October 1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/333,573
;; FILING DATE: 01 November 1994
;; ATTORNEY/AGENT INFORMATION:

;; NAME: Maria C.H. Lin
;; REGISTRATION NUMBER: 29,323
;; REFERENCE/DOCKET NUMBER: 1151-4118PC
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212)758-4800
;; TELEFAX: (212)751-6849
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 61 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
PCT-US95-13660-3

Query Match 91.0%; Score 142; DB 5; Length 61;
Best Local Similarity 92.9%; Pred. No. 1.8e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPDQVKGPGGQIVG 28
||||| |||||||||
Db 4 PKPQKTKRNTNRRPDQVKGPGGQIVG 31

RESULT 14
US-08-836-075A-10
;; Sequence 10, Application US/08836075A
;; Patent No. 6180768
;; GENERAL INFORMATION:
;; APPLICANT: MAERTENS, GEERT
;; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
;; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
;; TITLE OF INVENTION: AGENTS
;; NUMBER OF SEQUENCES: 207
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ARNOLD, WHITE & DURKEE
;; STREET: P.O. BOX 4433
;; CITY: HOUSTON
;; STATE: TEXAS
;; COUNTRY: USA
;; ZIP: 77210-4433
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Microsoft Word 6.0 / ASCII text output
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/836,075A
;; FILING DATE: 21 Apr 1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/EP95/04155
;; FILING DATE: 23 Oct 1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: EP 94870166.9
;; FILING DATE: 21 Oct 1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: EP 95870076.7
;; FILING DATE: 28 Jun 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KAMMERER, PATRICIA A.
;; REGISTRATION NUMBER: 29,775
;; REFERENCE/DOCKET NUMBER: INNS:004
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 74 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-836-075A-10

Query Match 91.0%; Score 142; DB 3; Length 74;
Best Local Similarity 92.9%; Pred. No. 2.2e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFPGGQIVG 28
||||| |||||||||
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

RESULT 15
US-08-635-886C-198
; Sequence 198, Application US/08635886C
; Patent No. 6555114
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2752-18
; CURRENT APPLICATION NUMBER: US/08/635,886C
; CURRENT FILING DATE: 1996-04-25
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 198
; LENGTH: 74
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-08-635-886C-198

Query Match 91.0%; Score 142; DB 4; Length 74;
Best Local Similarity 92.9%; Pred. No. 2.2e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFPGGQIVG 28
||||| |||||||||
Db 5 PKPQKTKRNTNRRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 11:23:50
Job time : 10.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on: August 7, 2003, 11:20:11 ; Search time 14.3636 Seconds
(without alignments)
231.506 Million cell updates/sec

Title: US-09-491-146A-26

Perfect score: 156
Sequence: 1 PKPQKPNRNTNRRPQDVKFPGGQIVG 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 451899 seqs, 118759770 residues

total number of bits satisfying chosen parameters:	451899
--	--------

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45
```

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	142	91.0	44	15	US-10-367-677-1	Sequence 1, Appl	
2	142	91.0	74	10	US-09-851-138-10	Sequence 10, Appl	
3	142	91.0	91	9	US-09-758-308-1	Sequence 1, Appl	
4	142	91.0	97	10	US-09-756-875-8	Sequence 8, Appl	
5	142	91.0	103	10	US-09-921-397-77	Sequence 77, Appl	
6	142	91.0	108	10	US-09-851-138-14	Sequence 14, Appl	
7	142	91.0	113	10	US-09-921-397-78	Sequence 78, Appl	
8	142	91.0	137	10	US-09-851-138-46	Sequence 46, Appl	
9	142	91.0	138	10	US-09-851-138-60	Sequence 60, Appl	
10	142	91.0	166	11	US-09-899-046-152	Sequence 152, App	
11	142	91.0	166	11	US-09-878-281-152	Sequence 42, Appl	
12	142	91.0	169	11	US-09-899-046-42	Sequence 44, Appl	
13	142	91.0	169	11	US-09-899-046-44	Sequence 42, Appl	
14	142	91.0	169	11	US-09-878-281-42	Sequence 44, Appl	
15	142	91.0	169	11	US-09-878-281-44	Sequence 44, Appl	

16	142	91.0	182	10	US-09-929-955-2	Sequence 2, Appl
17	142	91.0	182	11	US-10-104-966-2	Sequence 2, Appl
18	142	91.0	191	11	US-09-134-949-3	Sequence 3, Appl
19	142	91.0	318	10	US-09-851-138-76	Sequence 76, Appl
20	142	91.0	319	10	US-09-851-138-12	Sequence 12, Appl
21	142	91.0	319	10	US-09-851-138-18	Sequence 18, Appl
22	142	91.0	319	11	US-09-899-046-50	Sequence 50, Appl
23	142	91.0	319	11	US-09-899-046-52	Sequence 52, Appl
24	142	91.0	319	11	US-09-899-046-54	Sequence 54, Appl
25	142	91.0	319	11	US-09-899-046-144	Sequence 144, Appl
26	142	91.0	319	11	US-09-878-281-50	Sequence 50, Appl
27	142	91.0	319	11	US-09-878-281-52	Sequence 52, Appl
28	142	91.0	319	11	US-09-878-281-54	Sequence 54, Appl
29	142	91.0	319	11	US-09-878-281-144	Sequence 144, Appl
30	142	91.0	809	10	US-09-973-025-50	Sequence 50, Appl
31	142	91.0	809	11	US-09-899-303-50	Sequence 50, Appl
32	142	91.0	809	11	US-09-995-808-50	Sequence 50, Appl
33	142	91.0	809	11	US-09-995-860-50	Sequence 50, Appl
34	142	91.0	2894	10	US-09-941-611-23	Sequence 23, Appl
35	142	91.0	2894	15	US-10-044-995-23	Sequence 23, Appl
36	142	91.0	2895	15	US-10-259-275-40	Sequence 40, Appl
37	142	91.0	3011	9	US-09-742-659-4	Sequence 4, Appl
38	142	91.0	3011	10	US-09-952-572-9	Sequence 9, Appl
39	142	91.0	3011	10	US-09-929-955-1	Sequence 1, Appl
40	142	91.0	3011	10	US-09-747-419-20	Sequence 20, Appl
41	142	91.0	3011	11	US-09-891-894-3	Sequence 3, Appl
42	142	91.0	3011	14	US-10-104-966-1	Sequence 1, Appl
43	142	91.0	3011	15	US-10-259-275-20	Sequence 20, Appl
44	142	91.0	3012	10	US-09-238-076-2	Sequence 2, Appl
45	142	91.0	3012	11	US-09-995-937-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

```

US-10-367-677-1
; Sequence 1, Application US/10367677
; Publication No. US20030118604A1
;
; GENERAL INFORMATION:
;
; APPLICANT: JOLIVET, MICHEL
; APPLICANT: PENIN, FRANÇOIS
; APPLICANT: DALSON, PASCAL
; APPLICANT: LADAVIERE, LAURENT
; APPLICANT: LACOUX, XAVIER
;
; TITLE OF INVENTION: ANTIGENIC STRUCTURES
; TITLE OF INVENTION: COMPOUNDS, ANALOGS
; TITLE OF INVENTION: TREATING AN H
;
; FILE REFERENCE: 103959
;
; CURRENT APPLICATION NUMBER: US/10/
; CURRENT FILING DATE: 2003-02-19
;
; PRIOR APPLICATION NUMBER: US/09/38
; PRIOR FILING DATE: 1999-09-07
;
; PRIOR APPLICATION NUMBER: EARLIER
; PRIOR FILING DATE: EARLIER FILING
;
; NUMBER OF SEQ ID NOS: 11
;
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1

```

Query Match 91.0%; Score 142; DB 15; Length 44;

```
; Best Local Similarity 92.9%; Pred. No. 2.5e-12; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 2;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 4 PKPQKTRNTNRRPDQVKFPGGGQIVG 31

RESULT 2
US-09-851-138-10
; Sequence 10, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS: 207
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-09-851-138-10
Query Match 91.0%; Score 142; DB 10; Length 74;
Best Local Similarity 92.9%; Pred. No. 4.3e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 3
US-09-758-308-1
; Sequence 1, Application US/09758308
; Patent No. US20020090607A1
; GENERAL INFORMATION:
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS
; FILE REFERENCE: 14114.0349U2
; CURRENT APPLICATION NUMBER: US/09/758,308
```

```
; CURRENT FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: 60/092,339
; PRIOR FILING DATE: 1999-07-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Hepatitis C Virus
US-09-758-308-1

Query Match 91.0%; Score 142; DB 9; Length 91;
Best Local Similarity 92.9%; Pred. No. 5.3e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 4
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 91.0%; Score 142; DB 10; Length 97;
Best Local Similarity 92.9%; Pred. No. 5.7e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPQKPNRNTNRRPDQVKFPGGGQIVG 28
    ||||| ||||| ||||| ||||| |||||
Db 5 PKPQKTRNTNRRPDQVKFPGGGQIVG 32

RESULT 5
```



```
US-09-921-397-77
; Sequence 77, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
;   pathogenic strain of the hepatitis C virus and
;   applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-921-397-77

Query Match          91.0%; Score 142; DB 10; Length 103;
Best Local Similarity 92.9%; Pred. No. 6.1e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQRKPNRNTNRRPDVKFPGGGQIVG 28
Db 18 PKPQRKTRNTNRRPDVKFPGGGQIVG 45

RESULT 6
US-09-851-138-14
; Sequence 14, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
;   AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
;   AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-09-921-397-78
; Sequence 78, Application US/09921397
; Patent No. US20020151484A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
;   pathogenic strain of the hepatitis C virus and
;   applications thereof
; FILE REFERENCE: B4809A - JAZ
; CURRENT APPLICATION NUMBER: US/09/921.397
; CURRENT FILING DATE: 2001-08-02
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-921-397-78

Query Match          91.0%; Score 142; DB 10; Length 113;
Best Local Similarity 92.9%; Pred. No. 6.7e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQRKPNRNTNRRPDVKFPGGGQIVG 28
Db 5 PKPQRKTRNTNRRPDVKFPGGGQIVG 32

RESULT 7
US-09-851-138-46
; Sequence 46, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
;   AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
;   AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
```

;; FILING DATE: 21 Oct 1994
;; APPLICATION NUMBER: EP 95870076.7
;; FILING DATE: 28 Jun 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KAMMERER, PATRICIA A.
;; REGISTRATION NUMBER: 29,775
;; REFERENCE/DOCKET NUMBER: INNS:004
;; INFORMATION FOR SEQ ID NO: 46:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 137 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-851-138-46

Query Match 91.0%; Score 142; DB 10; Length 137;
Best Local Similarity 92.9%; Pred. No. 8.2e-12;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 PKPQRKPNRNTNRRPDVKFPGGGQIVG 28
5 PKPQRKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 9

US-09-851-138-60
; Sequence 60, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS

;; NUMBER OF SEQUENCES: 207
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ARNOLD, WHITE & DURKEE
;; STREET: P.O. BOX 4433
;; CITY: HOUSTON
;; STATE: TEXAS
;; COUNTRY: USA
;; ZIP: 77210-4433
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Microsoft Word 6.0 / ASCII text output
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/851,138
;; FILING DATE: 09-May-2001
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/836,075
;; FILING DATE: <Unknown>
;; APPLICATION NUMBER: EP 94870166.9
;; FILING DATE: 21 Oct 1994
;; APPLICATION NUMBER: EP 95870076.7
;; FILING DATE: 28 Jun 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KAMMERER, PATRICIA A.
;; REGISTRATION NUMBER: 29,775
;; REFERENCE/DOCKET NUMBER: INNS:004
;; INFORMATION FOR SEQ ID NO: 60:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 138 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 60:

US-09-851-138-60
Query Match 91.0%; Score 142; DB 10; Length 138;
Best Local Similarity 92.9%; Pred. No. 8.3e-12;

Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 PKPQRKPNRNTNRRPDVKFPGGGQIVG 28
||||| ||||||| ||||||| |||||||
Db 5 PKPQRKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 10

US-09-899-046-152
; Sequence 152, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:

;; TITLE OF INVENTION: New sequences of hepatitis C virus
;; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
;; NUMBER OF SEQUENCES: 270
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/899,046
;; FILING DATE:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/362,455
;; FILING DATE:

;; INFORMATION FOR SEQ ID NO: 152:
;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 166 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-899-046-152

Query Match 91.0%; Score 142; DB 11; Length 166;
Best Local Similarity 92.9%; Pred. No. 1e-11;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQRKPNRNTNRRPDVKFPGGGQIVG 28
||||| ||||||| ||||||| |||||||
Db 5 PKPQRKTKRNTNRRPDVKFPGGGQIVG 32

RESULT 11

US-09-878-281-152
; Sequence 152, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:

;; TITLE OF INVENTION: New sequences of hepatitis C virus
;; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
;; NUMBER OF SEQUENCES: 270
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/878,281
;; FILING DATE:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/362,455
;; FILING DATE:

;; INFORMATION FOR SEQ ID NO: 152:
;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 166 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-878-281-152

Query Match 91.0%; Score 142; DB 11; Length 166;

Query Match 91.0%; Score 142; DB 11; Length 169;
Best Local Similarity 92.9%; Pred. No. 1e-11; 2; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
||||| |||||||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 14

US-09-878-281-42
; Sequence 42, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-42

Query Match 91.0%; Score 142; DB 11; Length 169;
Best Local Similarity 92.9%; Pred. No. 1e-11; 2; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
||||| |||||||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 15

US-09-878-281-44
; Sequence 44, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-44

Best Local Similarity 92.9%; Pred. No. 1e-11; 2; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
||||| |||||||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 12

US-09-899-046-42
; Sequence 42, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-42

Query Match 91.0%; Score 142; DB 11; Length 169;
Best Local Similarity 92.9%; Pred. No. 1e-11; 2; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPQKPNRNTNRRPQDVKFGGQIVG 28
||||| |||||||||
Db 5 PKPQKTKRNTNRRPQDVKFGGQIVG 32

RESULT 13

US-09-899-046-44
; Sequence 44, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-44

Query Match 91.0%; Score 142; DB 11; Length 169;
 Best Local Similarity 92.9%; Pred. No. 1e-11;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PKPQKFNRTNRRPQDVKFPGGQIVG 28
 Db 5 PKPQKFNRTNRRPQDVKFPGGQIVG 32

Search completed: August 7, 2003, 12:01:13
 Job time : 15.3636 secs